Phonological Short-Term Memory and New Word Learning: Evidence from Paired-Associate and Hebb Repetition Paradigms

Elizabeth Ann Littlewood

Submitted for the degree of Doctorate of Philosophy

University of York

Department of Psychology

September 2007

Abstract

This thesis investigated the role of phonological short-term memory (PSTM) in the long-term learning of new phonological word-forms. Previous studies using the paired-associate paradigm have suggested that the learning of unfamiliar material is mediated by PSTM (e.g. Papagno & Vallar, 1992). The first aim was to replicate and extend this previous work. The second aim was to determine whether the Hebb repetition paradigm could provide an alternative method with which to investigate the role of PSTM in new word-form learning. Seven experiments were conducted to explore these aims.

Experiment 1 obtained phonological similarity effects for words and nonwords in an immediate serial recall task, confirming that the chosen manipulation of phonological similarity was adequate. Experiments 2 and 3 adopted the paired-associate task and replicated Papagno and Vallar (1992), thus extending their results to English participants and materials. Phonological similarity was shown to selectively disrupt the learning of nonword pairs. In contrast, some evidence was found to suggest that phonological similarity fails to affect the learning of word pairs. However, Experiment 3 showed that the detrimental effect of phonological similarity was restricted to an intermediate phase of learning. These findings suggest that PSTM mediates the learning of unfamiliar material, although the role of PSTM may change during the course of learning. Experiments 4 and 6 adopted the Hebb repetition task and generated patterns of results consistent with Papagno and Vallar (1992) and Experiments 2 and 3. Phonological similarity disrupted the learning of nonword sequences, but not the learning of word sequences. These findings suggest that PSTM mediates sequence learning for unfamiliar material, thereby providing initial evidence that the Hebb repetition paradigm may be a possible analogue of new word-form learning. The role of PSTM in nonword sequence learning could not be reliably assessed in Experiments 5 and 7 due to the absence of reliable Hebb Effects. Analyses of between-trial learning and forgetting rates using a Markov model revealed that phonological similarity had a negative impact on forgetting rates for nonwords in both paired-associate and Hebb repetition paradigms, suggesting that phonological representations of nonwords are particularly fragile. Finally, it is proposed that the paired-associate paradigm represents a closer analogue of new word-form learning than the Hebb repetition paradigm as it makes use of existing lexical-semantic information.

Contents

Chapter One: Phonological Short-Term Memory and the Long-Term Learning of New Word-Forms

Appendices

List of Tables

List of Figures

Acknowledgements

There are a number of people whom I would like to thank for their support and encouragement during my PhD. First and foremost, I would like to thank my supervisor, Graham Hitch. Graham inspired me to embark on a PhD and has provided me with invaluable support and guidance over the past four years. Graham has been a wonderful mentor; his enthusiasm for my research, his expertise and his unyielding patience and understanding, have made the production of this thesis possible. I hope we will work together again in the future.

I would also like to extend my gratitude to Alan Baddeley for being an active member of my Research Committee and for offering insightful research ideas, and to Susan Gathercole for her recent support and understanding. My thanks also go to Leesa Clarke for being the 'female native English speaker' in all my experiments and Philip Quinlan for statistics advice.

Special thanks go to my good friend, Steve King. He gave me the confidence to believe I could do a PhD and has been a great source of emotional support and understanding – thank you for believing in me. I would not have survived the highs and lows of writing up this thesis if it had not been for the continued support and reassuring words of Meesha Warmington – thank you so much. You have kept me sane. I hope I can do the same for you. My personal thanks go to a wonderful network of friends – Lynne Armitage, Joni Holmes, Siân Lindley, Hannah Nash, Michelle St. Clair and Helen Staniforth – all of whom have supported and motivated me enormously throughout my PhD. Thanks also go to everyone in C224 past and present, Paula Clarke, the Barnsley gang and Salsaholics! Special thanks go also to David Honan for his recent support, patience and understanding.

Finally, my love and thanks go to my Mum, Dad and sister Helen, for their unconditional love and support. You never once doubted me. Mum – thank you so much for always being on the other end of the phone with much needed words of encouragement – you never tired of my woes! I hope Grandma would have been proud of me too.

Mum, Dad, Helen – this is for you.

Author's Declaration

This thesis contains research conducted solely by the author under the supervision of Professor Graham Hitch.

This research was supported by a studentship from the Biotechnology and Biological Sciences Research Council.

The data reported in Chapter Three were presented at the 2004 BPS Cognitive Section Annual Conference:

Littlewood, E. & Hitch, G. (2004). *Does phonological similarity disrupt the learning of nonwords?* Poster session presented at the BPS Cognitive Section Annual Conference, University of Leeds, UK.

The data reported in Chapters Five and Six were presented at the 2005 BPS Cognitive Section Annual Conference and at the 2006 4th International Conference on Memory:

Littlewood, E. & Hitch, G. (2005). *Does phonological similarity play a role in nonword sequence learning?* Paper presented at the BPS Cognitive Section Annual Conference, University of Leeds, UK.

Littlewood, E. & Hitch, G. (2006). *The role of phonological short-term memory in new word-form learning: Evidence from a nonword sequence learning paradigm.* Paper presented at the Fourth International Conference on Memory, University of New South Wales, Sydney.

Chapter One: Phonological Short-Term Memory and the Long-Term Learning of New Word-Forms

1.1 Introduction

The aim of this chapter is to review the literature examining phonological short-term memory (PSTM) and its role in the long-term learning of new phonological word-forms. The chapter begins with a brief summary of why vocabulary acquisition is an important cognitive skill to investigate. It will then introduce the idea that PSTM may play an important role in the learning of new word-forms. The chapter is then split into two broad sections. The first of these introduces the Working Memory Model (WM model; Baddeley & Hitch, 1974) as the theoretical framework upon which much research conducted on PSTM and new word-form learning has been grounded. Particular emphasis is placed on reviewing the concept of the phonological loop component of the WM model. The second section reviews the evidence obtained from a wide variety of populations in support of the role of PSTM in language learning. Challenges to this viewpoint will also be described. The chapter concludes with an overview of the research questions this thesis aims to address.

1.2 The Importance of Vocabulary Acquisition

One of the most important cognitive skills a developing child needs to learn is the vocabulary of its native language. The capacity to learn new words underpins successful intellectual and educational development. Indeed, Sternberg (1987) suggests that vocabulary acquisition is "the single most important determinant of a child's eventual intellectual and educational attainments" (cited in Baddeley, Gathercole & Papagno, 1998). Furthermore, vocabulary knowledge is often used as an indicator of verbal intelligence (Gathercole & Baddeley, 1993) and is one of the skills assessed by standardised assessment tools, such as the British Picture Vocabulary Scales (BPVS; Dunn & Dunn, 1982).

Over the years, research has focused on the specific linguistic skills a child needs to develop in order to both produce and comprehend its native language. This research has examined skills such as how a child learns to correctly segment a perceptual stream of incoming speech into its constituent words (e.g. Jusczyk, Luce & Charles-Luce, 1994; Saffran, Newport & Aslin, 1996); how a child learns the concepts and meanings of the words they encounter (e.g. Carey, 1978; Markman, 1994; Markman, Wasow & Hansen, 2003; Markman & Wachtel, 1988); and how the syntactic structure of a language is acquired (e.g. Brown, 1973; Gleitman, 1993; Pinker, 1984; all cited in Baddeley et al., 1998). Whilst these linguistic skills obviously play a vital role in the language learning process, it is also important to consider how a child is able to *remember* the *sound patterns* of new words. Each word a child encounters for the first time is a new word and hence will be initially perceived as a novel phonological sequence. In this sense, vocabulary acquisition also involves the long-term learning of the representations of many different phonological sequences. But what mechanisms operate to enable these phonological sequences to become stable and permanent representations?

1.3 Is There a Role for Phonological Short-Term Memory in Vocabulary Acquisition?

Before introducing the idea that PSTM may have a role to play in vocabulary acquisition, it is important to define a number of terms adopted throughout this thesis and to highlight their usage in reference to other related terms. Firstly, the term 'short-term memory' (STM) is used to refer to the general capacity to store information (visual or verbal) over the short-term and is therefore not mentioned in connection to a specific model of STM, unless explicitly stated otherwise. Secondly, the term 'PSTM' is used to refer to the "well-documented capacities of both children and adults to retain sequences of *verbal* material over short periods of time" (Gathercole & Martin, 1996, p. 73, italics added). The use of this term is intended to reflect this capacity and as such does not offer any commitment to any particular model of STM. Finally, although the term 'phonological loop' is also used to refer to the capacity to maintain verbal (or phonological) material in STM, its use throughout this thesis refers directly to its status as a specific component of the theoretical WM model (Baddeley & Hitch, 1974).

As suggested in the previous section, vocabulary acquisition involves acquiring the phonological characteristics of new words as well as the semantic properties. As such, there has now been a considerable amount of research conducted into identifying the cognitive processes and mechanisms involved in the process of learning the sound patterns of new phonological sequences (Gathercole, Hitch, Service & Martin, 1997). As a result of this research, evidence has begun to accumulate which suggests that a relationship exists between vocabulary acquisition and the ability to temporarily store the representations of new phonological sequences (e.g. Gathercole & Baddeley, 1989a; Gathercole, Willis, Emslie and Baddeley, 1992; see also Baddeley et al., 1998, for a review). Such research has converged on the view that PSTM provides support for the "construction of more permanent representations of the phonological structure of new words" (Baddeley et al., 1998; p. 170). More specifically, it has been proposed that the phonological loop component of the WM model may be the mechanism which facilitates the long-term learning of novel phonological representations (Badddeley et al., 1998).

The following section therefore introduces the architecture of the WM model. Particular attention is given to describing the structure of the phonological loop as it is this component which is of particular relevance to this thesis.

1.4 The Working Memory Model

One of the most influential models of STM to date is the WM model developed by Baddeley and Hitch (1974) and later revised by Baddeley (1986, 2000a). In contrast to earlier theories which took the approach that STM reflected a unitary system (e.g. Atkinson & Shiffrin, 1968; Craik & Lockhart, 1972), the WM model posited that STM may serve as a more general working memory system which could be fractionated into three functionally specialised and independent subsystems (Baddeley, 1992). This fractionation was partly based on findings obtained with patients suffering neuropsychological damage; several case studies described patients with STM impairments but who showed no evidence of difficulties in every day cognitive functioning (e.g. Shallice & Warrington, 1970; Vallar & Papagno, 2002; Vallar & Shallice, 1990). If STM were a unitary system, then these patients would be expected to show severe problems with long-term learning and cognitive activities. The functional importance of such a fractionated memory system is further supported by its role in facilitating a wide range of complex cognitive abilities such as reasoning, learning and comprehension (e.g. Baddeley, 1986, 2003).

1.4.1 Architecture of the Working Memory Model

Baddeley and Hitch (1974) describe the WM model as a limited-capacity multi-component model comprising three components: the central executive, the visuospatial sketchpad and the phonological loop (see Figure 1.1). The central executive is regarded as a modality-independent attentional control system responsible for the coordination and integration of information from within its two subsidiary slave systems, the visuospatial sketchpad and the phonological loop. As such, the central executive is believed to be capable of performing a number of high-level functions (Gathercole, 1998). The two slave systems are designed for information storage with the visuospatial sketchpad supporting the retention of visual and spatial information (e.g. Logie, 1995) and the phonological loop supporting the maintenance of verbal or speech-based phonological information (e.g. Baddeley, 1986).

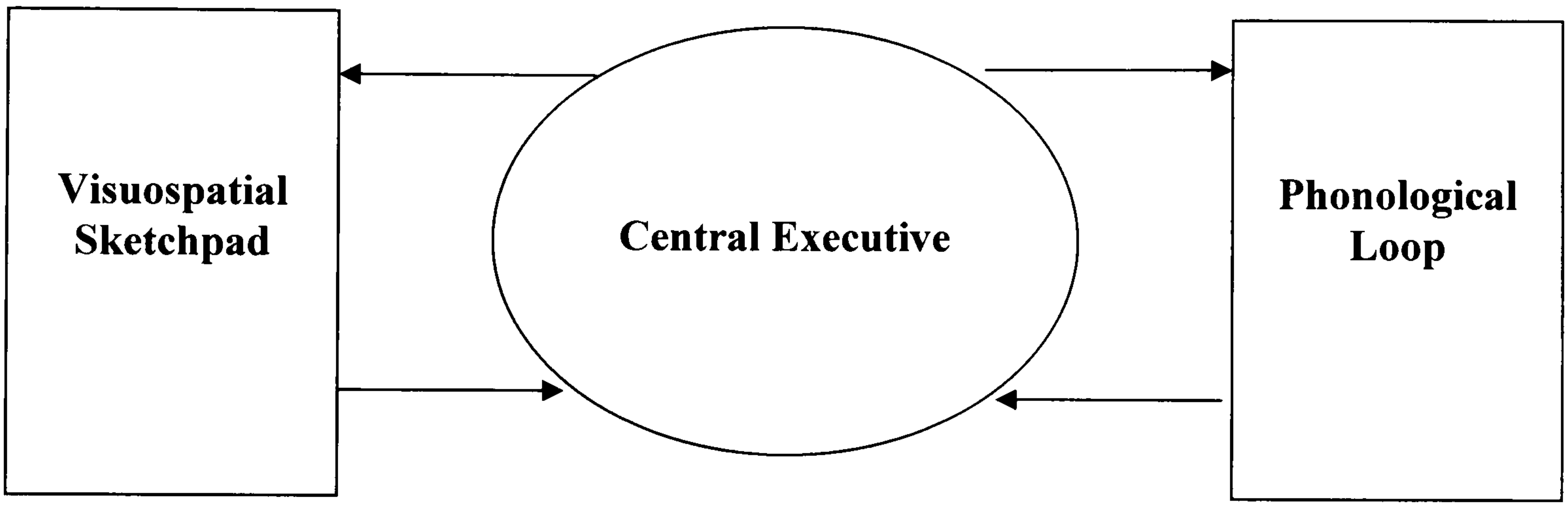


Figure 1.1: *Baddeley & Hitch's (1974) Working Memory Model (adapted from Baddeley, 2003)*

Studies investigating the WM model's two storage systems have converged on the view that they are functionally independent elements of STM. Such studies use a number of experimental techniques to assess verbal and visuospatial STM performance. The most common method of assessing verbal STM is the immediate serial recall (ISR) task. In this task, a sequence of items, such as digits, letters or unrelated words, is presented to participants aurally or visually. Immediately following sequence presentation, participants are required to recall the items back in their original serial order. Visuospatial STM is assessed by separate visual and spatial tasks on the basis that neuropsychological studies have shown double dissociations between visual and spatial span (e.g. Farah, Levine, Calvanio, 1988; Della Sala, Gray, Baddeley, Allamano & Wilson, 1999; see also Della Sala and Logie, 2002, for a neuropsychological review). Visual STM is assessed by pattern span (e.g. Visual Patterns Test; Della Sala, Gray, Baddeley & Wilson, 1997). In this task, participants are presented with checkerboard patterns in which some cells are filled. The checkerboard is removed and replaced by a blank checkerboard on which participants indicate which cells were originally filled. Spatial STM is often measured by the Corsi Blocks Test (Corsi, 1972). This task involves the experimenter pointing to a sequence of blocks arranged randomly on a board. Participants are then required to immediately repeat the sequence by pointing to the blocks in their original serial order.

Neuropsychological studies have shown dissociable impairments of the verbal (phonological loop) and visual (visuospatial sketchpad) components (e.g. Basso, Spinnler, Vallar & Zanobio, 1982; Farah, 1988; see also Vallar & Papagno, 2002, for a neuropsychological review) using verbal and visuospatial tasks. Patients with specific verbal STM deficits appear to have normal visuospatial function (e.g. Basso et al., 1982). Conversely, verbal STM is spared in some patients who suffer from marked deficits of visuospatial STM (e.g. Farah, 1988). Such studies are argued to provide evidence for the distinction between the verbal and visuospatial storage subsystems of STM (Gathercole, 1998). Consistent with this view, neuroimaging studies have reported that the phonological loop and visuospatial sketchpad are located in different brain regions. The phonological loop is argued to be located in the left temporoparietal regions involving Broca's area and the prefrontal cortex, whilst the visuospatial sketchpad is argued to reside in the right parietal and prefrontal areas (e.g. Baddeley, 2003; Gathercole, 1998; see also Smith, Jonides & Koeppe, 1996; Smith & Jonides, 1997, for neuroimaging reviews).

The phonological loop component of the WM model is arguably the subsystem which has been investigated in most detail to date (Baddeley, 2007). The following sections therefore describe the structure of the phonological loop and the evidence in support of its architecture. A brief overview of the development of the phonological loop in childhood will also be provided. The final section describes some of the challenges to the phonological loop.

1.4.1.1 The Phonological Loop

The phonological loop consists of two related but separable subcomponents: a passive phonological short-term store which holds speech-based information in a phonological code; and an active rehearsal process which maintains information in the phonological short-term store by a process of subvocal rehearsal (e.g. Baddeley, 1986). Subvocal rehearsal operates in real-time by refreshing decaying representations in the phonological short-term store in a time-based manner, thereby off-setting or preventing their degradation from STM (e.g. Baddeley, 1986).

The phonological store is assumed to be of limited capacity, with the number of items held in the store corresponding to the amount of information which can be rehearsed within approximately two seconds (Baddeley & Hitch, 1974; Baddeley, Thomson & Buchanan, 1975). Spoken stimuli are believed to gain automatic and direct access to the phonological short-term store, whereas visually presented stimuli, such as pictures or printed words, have to be converted into a phonological code before entering the phonological short-term store (e.g. Baddeley, 1986). This necessary conversion requires the operation of the subvocal rehearsal process; hence, if rehearsal is prevented, visual information cannot gain access to the phonological short-term store (e.g. Baddeley, 1986).

1.4.1.2 Evidence for the Phonological Loop

The concept of the phonological loop has proved capable of accounting for a number of key phenomena of PSTM (Baddeley, 2007; Gathercole, 1997). These refer to the effects of articulatory suppression, word length, phonological similarity and irrelevant speech. These have since become core characteristics that any model of PSTM needs to address (Gathercole, 1997). These phenomena will now be discussed.

Articulatory Suppression

The effects of articulatory suppression were first reported in a series of experiments conducted by Murray (1967, 1968). In these studies, participants perform an ISR task whilst simultaneously engaging in overt articulation of irrelevant material (e.g. *'the, the, the')* throughout the presentation of a memory list. The secondary task of articulatory suppression exerts a deleterious effect on ISR performance; recall levels are severely reduced compared to recall performance without suppression (e.g. Baddeley et al., 1975; Baddeley, Lewis & Vallar, 1984; Estes, 1973; Levy, 1971). However, although ISR performance is impaired, it is not completely obliterated, with participants accurately recalling several memory items.

The explanation for the effect of articulatory suppression on recall performance lies in the subvocal rehearsal component of the phonological loop. The procedure of articulating irrelevant speech-based information serves to interfere with the process of rehearsing the memory items. Articulatory suppression is assumed to prevent, or at least severely restrict, the use of the rehearsal mechanism as a means of refreshing the decaying memory traces of the items to-be-remembered (Baddeley, 1986). The result is a reduction in ISR performance. Articulatory suppression has an even greater effect on visually presented items as the translation of visual items into a phonological code is prevented (Baddeley et al., 1975).

Word Length Effect

The effect of word length on PSTM performance was initially investigated by Baddeley et al. (1975) and refers to the finding that ISR performance is better when sequences contain unrelated short words (e.g. *cat, pen, ball*) than when sequences contain unrelated long words (e.g. *banana, hippopotamus, university*). The word length effect (WLE) is remarkably robust. It occurs for word sequences presented auditorily or visually (e.g. Baddeley et al., 1975), for sequences containing nonwords instead of words (e.g. Hulme, Maughan & Brown, 1991), and also extends to children as young as 4 years of age when sequences are aurally presented (e.g. Hitch & Halliday, 1983; Hitch, Halliday, Dodd & Littler, 1989a; Hulme, Thomson, Muir & Lawrence, 1984).

Baddeley et al. (1975) argue that the WLE reflects the use of the subvocal rehearsal mechanism. They propose that fewer long words are recalled as these take longer to articulate and hence rehearse; that is, fewer long words will be refreshed during the process of subvocal rehearsal. As a result, long words are more likely to suffer from decay, leading to their rapid loss from the phonological short-term store. Further support for this account comes from Baddeley et al.'s (1975) observation of a linear relationship between rate of articulation and PSTM performance. Better performance was found for words with shorter spoken durations when controlling for phonological complexity (i.e. numbers of syllables and phonemes). Baddeley et al. (1975) concluded that the WLE is a product of the spoken duration of items and is not a consequence of the length (i.e. numbers of syllable and phonemes) of words *per se*.

The effect of articulatory suppression on the WLE also provides evidence for the operation of the subvocal rehearsal mechanism. The WLE remains when words are presented auditorily and articulatory suppression is required throughout the presentation phase (e.g. Baddeley et al., 1975). However, when articulatory suppression extends throughout presentation and recall, the WLE is eliminated (e.g. Baddeley et al., 1984). This suggests that the act of articulatory suppression throughout presentation and recall prevents participants from engaging in the time-based rehearsal process; as a result, the words are not refreshed and hence rapidly decay from the phonological short-term store (e.g. Baddeley, 1986). For visual presentation, articulatory suppression throughout presentation abolishes the WLE. In this instance, articulatory suppression prevents the necessary conversion of visual codes into phonological codes (e.g. Baddeley, 1986).

Phonological Similarity Effect

The phonological similarity effect (PSE) was first reported by Conrad and Hull (1964) and refers to the finding that sequences of phonologically similar sounding items (e.g. *B, G, D; man, cat, cap*) are harder to recall than sequences of phonologically distinct sounding items (e.g. *R, K, H; cow, bar, few*). The PSE is highly robust (e.g. Logie, Della Sala, Laiacona, Chalmers & Wynn, 1996) and is often taken to reflect the speech-based nature of the phonological loop. Indeed, Baddeley (2003) refers to the PSE as a "marker of the phonological loop" (p. 831).

The locus of the PSE has been argued to lie in the phonological short-term store. Memory items with similar sounding structures will leave memory traces which are less distinct than memory items with more distinct sounding structures. During decay of these memory traces over time, the similar sounding items will become less discriminable, leading to a greater degree of confusion between these items (Baddeley, 1986). The PSE emerges with both auditory and visual presentation with adults (e.g. Baddeley, 1966a; Conrad, 1964; Gathercole, Pickering, Hall & Peaker, 2001; Salamé & Baddeley, 1982, 1986; Wickelgren, 1965a); with very young children, provided the items are presented auditorily (e.g. Gathercole et al., 2001; Hulme, 1987); and with congenitally deaf children when the memory items are presented visually (e.g. Conrad, 1970).

The effect of phonological similarity also interacts with articulatory suppression. In contrast to the WLE, the PSE persists under articulatory suppression when the memory items are aurally presented, even when articulatory suppression continues throughout presentation and recall (Baddeley et al., 1984). This is assumed to reflect the automatic access of speech-based information to the phonological short-term store (Baddeley, 1986). However, with visual presentation, the PSE is removed under articulatory suppression (e.g. Levy, 1971; Peterson & Johnson, 1971), confirming that engaging in articulatory suppression prevents the translation of visual material into a phonological code, thereby preventing entry of this material into the phonological short-term store (Baddeley, 1986). The PSE has also been shown to disappear when the ISR task becomes too demanding (e.g. Larsen & Baddeley, 2003); under these circumstances it has been proposed that participants abandon their reliance on phonological coding and instead utilise alternative strategies such as semantic or visual coding (Baddeley, 2003).

Irrelevant Speech Effect

Salamé and Baddeley (1982) provided evidence that the ISR of a sequence of visually presented items was impaired by the simultaneous presentation of irrelevant speech in an unfamiliar language, which participants are instructed to ignore. It was claimed that the irrelevant speech automatically gains direct access to the phonological short-term store, thereby corrupting the memory traces of the memory items to-be-learned. In line with this view, articulatory suppression was shown to remove the irrelevant speech effect (ISE) by preventing the use of the rehearsal mechanism to convert the visual items into a phonological code.

The ISE was taken to reflect the speech-based nature of the phonological short-term store and has been found when the speech is in an unfamiliar language (e.g. Baddeley & Salamé, 1986; Colle & Welsh, 1976; Salamé & Baddeley, 1986) and when it consists of nonwords (Salamé & Baddeley, 1982). Salamé and Baddeley (1982) have also argued that the ISE does not operate at a lexical level; memory for digit sequences was found to be no more impaired by irrelevant digits than by words comprising the same phonemes as the digits (e.g. *tun, woo* instead of *one two*).

Summary

The evidence reviewed above suggests that the two-component phonological loop is able to account for the effects of articulatory suppression, word length, phonological similarity and irrelevant speech on PSTM performance (e.g. Baddeley, 1986; Baddeley & Salamé, 1986; Baddeley et al. 1975, 1984; Conrad & Hull, 1964; Levy, 1971; Murray, 1967, 1968; Peterson & Johnson, 1971; Salamé & Baddeley, 1983). However, it is important to acknowledge that there have since been numerous challenges to these interpretations. These will be discussed in section 1.4.1.4.

1.4.1.3 Development of the Phonological Loop

A large body of research has been conducted with the aim of investigating the developmental changes in the operation of the phonological loop (e.g. Halliday, Hitch, Lennon & Pettifer, 1990; Hitch & Halliday, 1983; Hitch, Halliday, Schaafstal & Schraagen, 1988; Hitch et al., 1989a; Hitch, Halliday & Littler, 1989b; Hitch, Halliday, Schaafstal & Heffernan, 1991; Hulme et al., 1984; Hulme & Tordoff, 1989; Nicolson, 1981; see Gathercole & Hitch, 1993 for a review). This research has suggested that although the phonological loop is present from the age of around 4 years and upwards, it is not fully functioning until about 7 years of age (Gathercole & Baddeley, 1993; Gathercole, 1998).

Many studies have investigated the emergence of the effects of word length and phonological similarity on ISR performance for children of differing ages (e.g. Halliday et al., 1990; Hitch & Halliday, 1983; Hitch et al., 1988, 1989a, 1991; Hulme & Tordoff, 1989). Collectively, these studies have shown that children over the age of 4 years are sensitive to both the WLE and the PSE. However, this sensitivity appears to interact with the modality of list presentation. Both effects have been reported with auditory presentation, demonstrating that children as young as 4 years of age are capable of utilising the subvocal rehearsal mechanism, albeit not as efficiently as older children or adults (Gathercole & Hitch, 1993). A different pattern emerges when memory items are presented visually in the form of pictures; children under 4 years do not show evidence of these effects (e.g. Hitch & Halliday, 1983; Hitch et al., 1991; Hulme et al., 1984; Hulme & Tordoff, 1989). However, these effects appear to emerge with visual presentation around 7 years of age (e.g. Halliday et al., 1990; Hitch et al., 1988, 1989a). These findings suggest that children under 7 years are unable to convert visual items into phonological codes via the subvocal rehearsal mechanism, despite being capable of utilising the rehearsal mechanism for aurally presented items (Gathercole & Baddeley, 1993; Gathercole & Hitch, 1993).

Evidence has also been reported for the role of subvocal rehearsal in the developmental increase in memory span - the mean number of words that can be recalled in the correct order - seen with age. Studies conducted by Nicolson (1981) and Hulme et al. (1984) found evidence in support of a linear relationship between speed of articulation and age. This suggested that the faster an individual's rate of rehearsal, the greater the number of items can be rehearsed before those items begin to decay from the phonological short-term store. Moreover, a study conducted by Hulme et al. (1991) showed that training in articulation of a foreign language both increased articulation rate and memory span in that language. This provided further support for the role of articulation rate or subvocal rehearsal in the development of memory span. A strong link between articulation rate and PSTM skills has also been reported in 2- and 3 year old children (Gathercole & Adams, 1993).

However, later studies questioned the use of subvocal rehearsal in children as young as 4 years. Gathercole, Adams and Hitch (1994) failed to find a significant correlation between rates of overt articulation and memory span for a sample of 4 year old children, although a significant relationship was reported in the same sample of children assessed one year later (e.g. Gathercole & Adams, 1994). Taking these findings together, these later studies converge on the view that young children cannot rehearse (Gathercole & Baddeley, 1993) and that use of the rehearsal mechanism does not emerge until 7 or 8 years of age (e.g. Flavell, Beach & Chinsky, 1966). However, this claim contradicts the earlier finding that 4 year old children are sensitive to the WLE when memory items are presented auditorily (e.g. Hitch & Halliday, 1983; Hulme et al., 1984). This raises the question of the validity of the claim that the WLE is attributable to the subvocal rehearsal process (e.g. Baddeley et al., 1975). Indeed, a number of alternative interpretations have been forwarded to account for the WLE in terms other than subvocal rehearsal (e.g. Caplan, Rochon & Waters, 1992; Caplan & Waters, 1994).

1.4.1.4 Challenges of the Phonological Loop

Despite the abundance of evidence reviewed supporting a phonological loop explanation of the empirical evidence obtained from PSTM tasks, numerous challenges have been forged offering alternative explanations for these PSTM phenomena (e.g. Caplan et al., 1992; Caplan & Waters, 1994; Cowan, Nugent, Elliot & Geer, 2000; Gupta & MacWhinney, 1995). Indeed, some of these challenges argue against the concept of PSTM itself (e.g. Jones & Macken, 1993; Jones, Macken & Nicholls, 2004; Jones, Hughes & Macken, 2006). These alternative accounts pose a number of difficulties for the concept of the phonological loop.

Word Length Effect

Perhaps the effect that has received the most controversy to date is the WLE (Baddeley, 2007). Numerous studies have been conducted demonstrating alternative accounts of the locus of the WLE (e.g. Avons, Wright & Pammer, 1994; Brown & Hulme, 1995; Caplan et al., 1992; Caplan & Waters, 1994; Cowan et al., 1992; Cowan, Wood, Nugent & Treisman, 1997; Cowan et al., 2000; Henry, 1991; Longoni, Richardson & Aiello, 1993; Lovatt, Avons & Masterson, 2000; 2002; Service, 1998; 2000). These challenges stem around two main lines of argument: one explains the WLE in terms of phonological complexity – the number of syllables and phonemes within an item - and the other in terms of output delay at recall.

A number of researchers have argued that the WLE stems from speech planning processes rather than the articulatory duration of words (e.g. Caplan et al., 1992; Caplan & Waters, 1994; Service, 1998). Caplan et al. (1992) suggest that speech planning times are influenced by the phonological complexity of words. The authors report better ISR performance for long compared to short words when the two sets of words were equated for phonological complexity. However, Baddeley and Andrade (1994) have criticised Caplan et al.'s (1992) study, claiming that articulatory durations were not accurately measured in their study and that they failed to equate the phonological similarity of the word sets. Despite this criticism, other studies have reported no effect of word length when controlling for phonological complexity (e.g. Caplan & Waters, 1994; Lovatt et al., 2000).

Other researchers point to the spoken recall process as the source of the WLE (e.g. Avons et al., 1994; Brown & Hulme, 1995; Cowan et al., 1992, 1997, 2000). Cowan et al. (1992) presented participants with memory lists containing both long and short words; in some lists long words followed short words, and in other lists short words followed long words. Recall performance was poorer for those lists containing long words at the beginning, resulting in the later (short) words decaying from PSTM. The authors suggest that the effects of word length occur due to the greater opportunity for decay of long compared to short items during the spoken recall phase. Brown and Hulme (1995) developed a simple trace decay model which was capable of simulating the effects of word length. The model did not incorporate any form of a rehearsal mechanism, instead basing the effects of word length on output interference at recall. However, Baddeley, Chincotta, Stafford & Turk (2002) found evidence for WLEs when controlling for output delay by using a serial recognition task – in this task participants are presented with a sequence of items; this sequence is followed by an identical sequence or one in which the position of two adjacent items has been switched. Participants simply have to indicate whether the two sequences were presented in the same order. The authors suggest that "the word length effect can influence retention through both rehearsal and output factors, as proposed by the phonological loop hypothesis" (p. 353).

In response to the challenges of a rehearsal explanation of the WLE (e.g. Caplan et al., 1992; Caplan & Water, 1994; Cowan et al., 1997, 2000; Lovatt et al., 2000, 2002; Service, 1998; 2000), a recent meticulous study conducted by Mueller, Seymour, Kieras and Meyer (2003) claim to provide evidence which refutes these alternative explanations. The authors argue that previous conflicts stem from "less than ideal measurement and control of factors such as of articulatory duration, phonological similarity and phonological complexity" (p. 1353). Mueller et al. (2003) developed a number of 'theoretically principled methods' designed to accurately quantify phonological dissimilarity and articulatory duration. Following these methods, the authors provided results which successfully accounted for the WLE in terms of spoken duration, thereby dispelling previous criticisms of a rehearsal explanation of the WLE.

A further source of evidence against a rehearsal explanation of the WLE comes from studies conducted with individuals with either acquired or congenital anarthria - the inability to produce discriminable speech sounds (Gathercole & Baddeley, 1993). These individuals show the normal disruptive effect of word length on PSTM performance (e.g. Baddeley & Wilson, 1985; Bishop & Robson, 1989), although a more recent study failed to find WLEs with anarthic individuals (e.g. Carlesimo, Galloni, Bonanni & Sabbadini, 2006).

Phonological Similarity Effect

The effect of phonological similarity on ISR performance has more recently been subject to a re-evaluation. A study by Jones et al. (2004) presented participants with auditorily or visually presented sequences of seven phonologically similar or distinct letters either with or without articulatory suppression. The expected interaction between phonological similarity, articulatory suppression and presentation modality was obtained (e.g. Baddeley et al., 1984; Levy, 1971; Peterson & Johnson, 1971). However, Jones et al. (2004) attributed the interaction to the emergence of a recency effect for distinct letters under articulatory suppression with auditory presentation only. They interpreted this in terms of the modality effect in which recency is enhanced with auditory presentation compared with visual presentation (e.g. Crowder & Morton, 1969). Moreover, Jones et al. (2004) succeeded in eliminating this interaction when adding a spoken suffix; that is, a redundant item to the end of the memory list. Jones et al. (2004) concluded that the phonological store is in fact not phonological at all, and is instead more accurately described as incorporating a “combination of auditory-perceptual and output planning mechanisms” (p. 656).

A later study by Jones et al. (2006) addressed the possibility that their previous finding (Jones et al., 2004) may be a consequence of using 7-item memory lists, given that the PSE has been shown to disappear when sequence length increases (e.g. Larsen & Baddeley, 2003). In their subsequent study, Jones et al. (2006) auditorily presented participants with either (i) 5-item memory lists with the addition of an extra redundant item presented at the start (prefix) and end (suffix) of each list; or (ii) 5-item memory lists without a prefix or a suffix. Both conditions were tested with and without articulatory suppression. This revealed a three-way interaction: the effect of phonological similarity was confined to a recency effect for distinct letters under articulatory suppression, with this recency effect being removed by the presence of prefix and a suffix. Jones et al. (2006) argued that this further supported their auditory-perceptual account of the PSE. In a rebuttal of these findings, Baddeley and Larsen (2007a) present evidence that, with 6-item memory lists, the interaction between phonological similarity, articulatory suppression and presentation modality emerges. Moreover, they show that the PSE is present at all portions of the serial position curve with auditory presentation, but is completely eliminated with visual presentation. Despite this latter finding, the debate regarding the extent to which the phonological short-term store is phonological in nature is still ongoing (e.g. Baddeley & Larsen, 2007b; Jones, Hughes & Macken, 2007).

Finally, a study conducted by Fallon, Groves and Tehan (1999) has reported evidence for the PSE for visually presented material under conditions of articulatory suppression. This result questions the extent to which articulatory suppression prevents the conversion of a visual code into a phonological code. The authors also argue that the way in which phonological similarity is operationalised can affect PSTM performance, suggesting that memory lists containing rhyming similar words facilitates ISR performance to a greater extent compared to memory lists containing nonrhyming similar words.

Irrelevant Speech Effect

The claim that the ISE reflects the speech-based nature of the phonological loop has since been refuted by Jones and colleagues (Jones, 1993; Jones, Beaman & Macken, 1996; Jones & Macken, 1993; Jones et al. 2004). Jones and Macken (1993) demonstrated that a simple fluctuating pure tone disrupted memory for visually presented items, whilst a single repeated tone failed to reliably disrupt memory. Jones (1993) proposed the changing state hypothesis to explain this result, claiming that it is the changing state of irrelevant sounds that impairs the storage of verbal material by interfering with the processing of serial order of memory list items. Jones and Macken's (1993) finding that irrelevant sounds disrupt memory was taken as evidence against the speech-based nature of the ISE. Furthermore, it has also been shown that the degree of phonological similarity between the memory list items and the irrelevant material does not influence the ISE; that is, irrelevant material that is phonologically similar sounding to memory list items causes no more disruption than phonologically distinct sounding irrelevant material (e.g. Jones & Macken, 1995; Larsen, Baddeley & Andrade, 2000; Salamé & Baddeley, 1986).

Serial Order and Long-Term Memory

In addition to the challenges of a phonological loop explanation of the key PSTM phenomena, there are two further problems which the phonological loop fails to address. The first of these refers to the absence of a mechanism responsible for the storage and processing of serial order. This is surprising considering that much of the research in the area of PSTM is based on the ISR task, in which the serial order of a sequence of items needs to be encoded, stored and retrieved. The phonological loop fails to specify how this order information is represented. A second important limitation of the phonological loop concerns its failure to offer an explanation as to how long-term learning occurs. There is no specified mechanism which explains the crucial interaction between PSTM and LTM. How does information entering the phonological loop culminate in permanent and stable long-term representations? How is information from LTM integrated into PSTM?

However, attempts have since been made to rectify these limitations. The development of a number of computational models of PSTM have focussed on implementing how serial order is represented in PSTM (e.g. Brown, Preece & Hulme, 2000; Burgess & Hitch, 1992, 1999, 2006; Gupta & MacWhinney, 1997; Henson, 1998; Lewandowsky & Murdock, 1989; Page & Norris, 1998). Furthermore, in an attempt to address the interaction between PSTM and LTM, Baddeley (2000a) has recently postulated a fourth component of the WM model. This component has been termed the episodic buffer. Baddeley (2000a) proposes the episodic buffer represents an interface between the three subsystems of the WM model and LTM. It is argued to be of limited capacity, temporary in nature and capable of representing information in a multi-dimensional code. Thus, the episodic buffer may represent a binding mechanism whereby information from different components of the WM model are combined and temporarily stored (Baddeley, 2007). The implementation of this episodic buffer may provide the necessary mechanism by which knowledge from LTM is integrated with PSTM.

Summary

This section has presented a considerable amount of research suggesting that the concept of the phonological loop is inadequate in its explanation of a number of key PSTM phenomena. Many of these challenges have provided convincing accounts of the WLE in terms of processes and mechanisms other than the subvocal rehearsal mechanism; these include phonological complexity and output delay at recall (e.g. Brown & Hulme, 1995; Caplan et al., 1992; Caplan & Waters, 1994; Cowan et al., 1992, 1997, 2000). However, the majority of these alternative accounts attempt to re-evaluate and extend the role of the phonological loop rather than reject it. For example, explanations of the WLE in terms of output delay do not necessary rule out the role of the phonological loop. Indeed, Baddeley et al.'s (2002) findings suggest that rehearsal and output delay factors contribute to the WLE.

Criticisms of the PSE in terms of the phonological loop have only recently been made and the debate as to whether the PSE represents a phonological short-term memory system or an auditory-perceptual system continues (e.g. Baddeley & Larsen, 2007a, 2007b; Jones et al., 2004, 2006, 2007). This debate is based primarily on alternative interpretations of the critical interaction between phonological similarity, articulatory suppression and presentation modality (Baddeley, 2007). However, as suggested by Baddeley and Larsen (2007a, 2007b), differences in this interaction between studies can feasibly be explained in terms of memory list length; using 7-item memory lists (e.g. Jones et al., 2004), or 5-item memory lists plus a prefix and suffix (e.g. Jones et al., 2006), can lead to participants abandoning any reliance on the phonological loop. Finally, the finding that the effect of irrelevant material on PSTM performance is disrupted by both sounds and speech appears to provide a problem for the speech-based nature of the phonological short-term store (e.g. Jones & Macken, 1993).

Arguably, some of the alternative interpretations of key PSTM phenomena may be attributable to inconsistencies across studies in the stimuli used and how well specific experimental design aspects are controlled. For example, articulatory duration and phonological similarity appear to be defined and controlled in different ways across studies. Indeed, this is suggested by the study conducted by Mueller et al. (2003).

Furthermore, the use of supraspan memory lists may serve to produce differential results across studies (e.g. Larsen & Baddeley, 2003).

This section has also acknowledged the limitations regarding the absence of mechanisms for (i) serial order and (ii) the integration of information from LTM. However, efforts have now been made to address these limitations in the form of computational models (e.g. Brown et al., 2000; Burgess & Hitch, 1999, 2006; Page & Norris, 1998) and the inclusion of the episodic buffer in the WM model (Baddeley, 2000a).

1.5 Influence of Existing Language Knowledge on Phonological Short-Term Memory

As previously described, the phonological loop provides no mechanism dedicated to the integration of information from LTM into PSTM. In this respect, the phonological loop assumes no influence of existing language knowledge on PSTM (but see Baddeley, 2000a). However, evidence has been documented suggesting that existing language knowledge serves to affect PSTM performance (e.g. Hulme et al., 1991; Hulme, Roodenrys, Brown & Mercer, 1995; Hulme et al., 1997; Gathercole, Willis, Emslie & Baddeley, 1991a; Gathercole, 1995; Gathercole, Frankish, Pickering & Peaker, 1999; Greg, Freedman & Smith, 1989; Roodenrys, Hulme, Alban, Ellis & Brown, 1994; Thorn & Frankish, 2005). Indeed, this evidence may suggest that there exists an interactive relationship between PSTM and long-term knowledge of the structure of language. If such a relationship exists, then it may also be plausible to suggest that PSTM influences the language learning process. The current section reviews the influence of stored language knowledge on PSTM performance. The remainder of the chapter is devoted to reviewing evidence in support of the idea that PSTM influences the language learning process.

Word Frequency Effect

Early evidence for long-term lexical effects in PSTM refers to the effect of word frequency. Watkins (1977) found that PSTM performance was higher when high-frequency words were presented in the first half of a memory list than when they were presented in the second half of a memory list. This was taken to reflect the influence of LTM for words at the beginning of a list, with words at the end of the list reflecting the influence of PSTM. However, Wright (1979) later argued that Watkins's (1977) finding was attributable to the longer duration of the low-frequency words. More recent research has shown evidence for word frequency effects that were not accounted for by differences in articulation rate (e.g. Roodenrys et al., 1994). Furthermore, the word frequency effect has been shown to survive the effect of articulatory suppression, thereby suggesting that the effect is not solely reliant on the phonological loop (e.g. Gregg, Freedman & Smith, 1989; Tehan & Humphreys, 1988). Hulme et al. (1997) also demonstrated word frequency effects on ISR performance when controlling for articulation rate. Moreover, in contrast to Watkins's (1979) finding, the effect increased as serial position increased, confirming that the effect is not due to retrieval from LTM of items early in the memory list. Hulme et al. (1997) conclude that their findings reflect differences in the accessibility of phonological representations of items in LTM, with high-frequency items providing more efficient support during a 'redintegrative' process, whereby "long-term memory representations are used to reconstruct partially decayed short-term phonological traces" (Hulme et al., 1991, p. 535).

Lexicality Effect

Hulme et al. (1991) observed that ISR performance was superior for lists of words compared to lists of nonwords. Moreover, this difference was found to be independent of articulation rate as indexed by equivalent WLEs for both types of memory lists. Hulme et al. (1991) attributed this lexicality effect to the absence of existing representations of the phonological forms of nonwords in LTM. In line with this, further evidence suggests that increasing participants' familiarity of nonwords results in improvements in memory span for those items (e.g. Hulme et al., 1991, 1995). Hulme et al. (1995) suggest this lexical contribution from LTM reflects the process of redintegration.

Wordlikeness Effect

Given the finding that memory for unfamiliar nonwords reflects memory for temporary phonological representations, as opposed to the activation of stored knowledge when lexical items are used (e.g. Hulme et al., 1991, 1995), the use of nonwords in PSTM tasks led to the claim that nonword memory measures provide 'purer' measures of PSTM ability (Gathercole & Martin, 1996). One such measure of PSTM capacity which is commonly used is the nonword repetition task. This task provides a measure of the accuracy with which an individual can repeat back an unfamiliar spoken form such as *woogalamic* (Baddeley et al., 1998). The importance of this task is reflected by the development of the Children's Test of Nonword Repetition (CNRep; Gathercole, Willis, Baddeley & Emslie, 1994).

However, the view that nonword memory measures provide a pure measure of PSTM has since been regarded as oversimplistic in light of evidence suggesting that the phonological structure of the language can also influence PSTM performance. Nonwords rated as being high in 'wordlikeness' (i.e. the degree to which a nonword would pass for a real word in the English language), such as '*defermication*', have been found to be better recalled than nonwords rated as low in wordlikeness, such as '*loddernapish*' (Gathercole et al., 1991a; Gathercole, Frankish et al., 1999; Grant et al., 1997; Vitevitch, Luce, Charles-Luce & Kemmerer, 1997; von Bon & van der Pijl, 1997). Furthermore, Gathercole et al. (1991a) reported a significant relationship between nonword repetition accuracy and wordlikeness ratings in a sample of 4- and 5 year old children. A later study by Gathercole (1995) replicated these findings when equating nonwords of high and low wordlikeness for phonological information (i.e. number of phonemes and syllables). Similar findings have also been reported with adults learning pairs of nonwords differing in wordlikeness (Gathercole, Martin & Hitch, 1996, cited in Gathercole & Martin, 1996). Taking these findings together, the effect of wordlikeness on PSTM performance indicates that PSTM can be strongly influenced by the phonological structure of language, as well as the lexical nature of language.

Phonotactic Frequency and Neighbourhood Size Effects

Recent research has argued that PSTM for nonwords is not only influenced by lexical knowledge from LTM (e.g. Roodenrys & Hinton, 2002; Roodenrys, Hulme, Lethbridge, Hinton & Nimmo, 2002), but also by sublexical LTM knowledge concerning the phonotactic frequencies of the language (e.g. Gathercole, Frankish et al., 1999; Thorn, Gathercole & Frankish, 2005; Thorn & Frankish, 2005). Phonotactic frequency refers to the statistical frequency of characteristic phoneme combinations in a language (Gathercole, Frankish et al., 1999; Roodenrys & Hinton, 2002; Thorn &

Frankish, 2005) and is typically assessed by comparing nonwords with high and low biphone frequencies (i.e. the frequency of occurrence of phoneme pairs in a language; for example, the biphone frequency of a consonant-vowel-consonant (CVC) nonword consists of two sets of phoneme pairs: CV_ and _VC). Gathercole, Frankish et al. (1999) reported superior recall for nonwords composed of high biphone-frequencies compared with nonwords composed of low biphone frequencies, suggesting an influence from LTM of the phonological properties of a language.

However, Roodenrys and Hinton (2002) reported no difference in recall performance for high and low biphone frequency nonword sets when these were matched for lexical neighbourhood size. Neighbourhood size has been defined as the number of words that "differ from the target word by the substitution of a single phoneme at any position" (Roodenrys et al., 2002; p. 1021). Instead, the authors found better recall performance for nonwords with large compared with small neighbourhoods when the two sets of nonwords were equated for biphone frequency. On the basis of these findings, they concluded that nonword recall is influenced by lexical rather than phonological knowledge.

However, a later study by Thorn and Frankish (2005) highlighted the high correlation between biphone frequency and neighbourhood size (e.g. Vitevitch, Luce, Pisoni & Auer, 1999), suggesting a possible confound in the Roodenrys and Hinton (2002) study. Furthermore, they criticised the measurement of neighbourhood size adopted by Roodenrys and Hinton (2002), pointing out that they only took into account CV_ and _VC neighbours, disregarding the C_C neighbours. Incorporating the number of C_C neighbours into the neighbourhood size calculation, Thorn and Frankish (2005) found a beneficial effect of neighbourhood size for nonword recall when the two sets of nonwords were equated for biphone frequency, in line with the findings from Roodenrys and Hinton (2002). However, Thorn and Frankish (2005) also found a significant influence on nonword recall performance of biphone frequency when the two sets of nonwords were matched on the neighbourhood size. Given these findings, the authors concluded that long-term knowledge contributions to nonword recall are based on both lexical and phonotactic knowledge of a language.

Summary

This section has reviewed a number of studies which provide convincing evidence that PSTM performance is influenced by stored knowledge of the language (e.g. Gathercole et al., 1991a, 1995; Gathercole, Frankish et al., 1999; Hulme et al., 1991, 1995, 1997; Roodenrys & Hinton, 2002; Thorn & Frankish, 2005). Furthermore, this research has shown that PSTM is influenced by both lexical and sublexical knowledge from LTM. PSTM performance is better for words than nonwords (e.g. Hulme et al., 1991, 1995), and high- than low frequency words (Hulme et al., 1997). PSTM is also influenced by factors such as the wordlikeness of nonwords (e.g. Gathercole, 1991a, 1995) and the phonotactic structure of the language (e.g. Roodenrys & Hinton, 2002; Thorn & Frankish, 2005). These findings suggest that PSTM performance "reflects a combination of temporary phonological storage and long-term knowledge concerning the structure of the language" (Gathercole, 1997, p. 25). Finally, this research highlights the importance of identifying a mechanism which is responsible for integrating information from LTM with PSTM. Indeed, as mentioned previously, Baddeley's (2000a) proposal of an episodic buffer may provide a suitable system for this role.

1.6 Interim Summary

The first section of this chapter introduced the WM model, with particular emphasis on the concept of the phonological loop component of this model. It has been shown that the phonological loop can successfully account for a number of well-documented PSTM phenomena, namely the effects of word length, phonological similarity, articulatory suppression and irrelevant speech (e.g. Baddeley et al., 1975; Conrad, 1964; Murray, 1967, 1968; Salamé & Baddeley, 1982). These effects have been explained in terms of the operation of the phonological short-term store and/or the subvocal rehearsal mechanism (e.g. Baddeley, 1986). The subvocal rehearsal mechanism has also been heavily implicated in the development of the phonological loop in young children (e.g. Hitch et al., 1988, 1989a, 1991; Hulme et al., 1984). Studies suggest that although this component is present in young children, they are not capable of utilising it efficiently until around 7 or 8 years of age (e.g. Gathercole, Adams et al., 1994).

Despite the evidence in support of a phonological loop explanation of key PSTM phenomena, several challenges have been put forward suggesting alternative processes and mechanisms to account for these phenomena. The WLE has been explained in terms of phonological complexity (e.g. Caplan et al., 1992; Caplan & Waters, 1994) and also output delay at recall (e.g. Brown & Hulme, 1995; Cowan et al., 1992, 1997, 2000). It has been argued that the PSE does not reflect phonological coding with PSTM (e.g. Jones et al., 2004, 2006). Indeed, Jones and colleagues propose that the phonological loop does not exist and that the PSE is purely a sensory effect reflecting auditory-perceptual mechanisms. Jones and colleagues (Jones, 1993; Jones & Macken, 1993; Jones et al., 1996) have also shown that the ISE extends to fluctuating sounds, proposing the changing state hypothesis to explain this finding. It is argued, however, that some of these alternative explanations may be attributed to inconsistencies in experimental design issues across studies (e.g. Baddeley & Andrade, 1994; Larsen & Baddeley, 2003; Meuller et al., 2003).

Two further limitations of the phonological loop have also been highlighted; namely, the lack of mechanisms responsible for serial order and for the integration of LTM into PSTM. Finally, evidence to suggest an interactive relationship between LTM and PSTM was reviewed. This evidence suggests that PSTM performance is influenced by stored knowledge of the language; this includes long-term knowledge about the lexical (i.e. word frequency, lexicality) and sublexical (e.g. wordlikeness, phonotactic frequency) characteristics of the language.

Despite these theoretical and empirical challenges, the phonological loop still remains an influential feature of PSTM. Indeed, over the past 20 years, researchers have been investigating the influence of PSTM in the important process of vocabulary acquisition. This research has converged on the claim that PSTM has evolved to play a vital role in the long-term learning of new phonological word-forms (e.g. Baddeley et al., 1998).

1.7 Phonological Short-Term Memory and New Word-Form Learning

The remainder of this chapter focuses on reviewing the extensive literature exploring the claim that PSTM contributes to the long-term learning of the phonological representations of new word-forms (e.g. Baddeley et al., 1998). The crux of this research is based on the view that a temporary representation of a novel phonological word-form is required in order for that representation to be learned. Indeed, in a comprehensive paper reviewing the evidence in support of this view, Baddeley et al. (1998) propose that the "primary function of the phonological loop is to provide temporary storage of unfamiliar phonological forms while more permanent memory representations are being constructed" (Baddeley et al., 1998; p. 159). This claim will be henceforth referred to as the PSTM hypothesis.

1.7.1 Support for the Phonological Short-Term Memory Hypothesis

The following sections review a large body of evidence examining the support for the PSTM hypotheis. These investigations cover a diverse range of population samples including; neuropsychological patients, normal adults, typically-developing children and children with specific language impairment (SLI); and utilise various methodologies including; single case studies, individual differences and experimental approaches. The final section discusses a number of alternative interpretations concerning phonological contributions to vocabulary acquisition.

1.7.1.1 Evidence from Neuropsychological Case Studies

The earliest evidence for the importance of PSTM in new word-form learning was provided by neuropsychological patients with acquired impairments of PSTM (e.g. Shallice & Warrington, 1970; Vallar & Baddeley, 1984a, 1984b). Although many of these patients suffered from a severely impaired PSTM, they were able to conduct normal lives and had only limited problems with everyday cognitive activities. One such patient, PV, has been extensively studied, providing the impetus for the subsequent surge in studies investigating PSTM and language learning (e.g. Baddeley, Papagno & Vallar, 1988; Vallar & Baddeley, 1984a, 1984b).

Patient PV was an Italian woman who had suffered a left hemisphere stroke. Although she performed within the normal range on a variety of cognitive abilities (e.g. speech fluency, articulation rate, phonological processing), her auditory-verbal span was dramatically reduced at only 2 to 3 items (Vallar & Baddeley, 1984b).

Following intensive investigation, Vallar and Baddeley (1984a) proposed that, although PV's phonological short-term store was functioning, as evidenced by her sensitivity to the PSE with auditory presentation, it was defective, resulting in her reduced auditory memory span. PV also failed to utilise the subvocal rehearsal mechanism to refresh decaying phonological representations in the phonological short-term store, or to convert visual information into a phonological code.

A subsequent study conducted with PV established the first direct evidence of a link between PSTM and language learning. Baddeley et al. (1988) examined PV's ability to learn the phonological forms of new words. Consistent with her reduced auditory memory span, PV presented a low memory span for nonwords. Of particular interest was PV's performance on a paired-associate learning task. In this task, individuals are initially presented with cue-target pairs of stimuli, later recalling the target stimulus in response to the presentation of the associated cue stimulus. PV was able to learn pairs of auditorily presented unrelated meaningful words in her native language to a comparable degree to that of a group of age- and education-matched controls. However, PV was completely unable to learn words paired with nonwords derived from Russian, a language PV was not familiar with (e.g. *Rosa-Svieti*). Over the course of ten trials, PV failed to learn any of the word-nonword pairs, whereas all of the controls correctly learned all word-nonword pairs by the final trial. PV's learning of word-nonword pairs did improve slightly with visual presentation, which may reflect the use of a visual coding strategy.

Baddeley et al.'s (1988) study confirmed that the phonological loop, particularly the phonological short-term store, was important for the long-term learning of the phonological forms of unfamiliar words. PV's unimpaired capacity to learn word-word pairs was presumed to reflect the use of non-phonological learning codes, such as semantic codes, given that words have existing lexical-semantic representations in LTM. However, in the absence of such existing representations for nonwords, coupled with impaired phonological storage, PV failed to construct stable phonological representations of the nonwords in LTM, thereby resulting in an inability to learn the word-nonword pairs.

PV's pattern of neuropsychological deficits has since been replicated. Trojano and Grossi (1995) report the case study of an Italian patient, SC, who had an acquired impairment of the phonological loop as a result of damage to the left temporal and parietal brain areas. SC had an auditory memory span of less than three but performed normally on a range of visuospatial tasks. Furthermore, SC showed the PSE with auditory but not visual presentation, and failed to show the WLE with either presentation modality. On this basis, SC was assumed to have both a defective phonological short-term store and an inability or unwillingness to utilise the rehearsal mechanism. In line with PV, SC was able to learn word-word pairs in his native language, but failed to learn word-nonword pairs, although his performance improved with visual presentation. However, unlike the Baddeley et al. (1988) study, SC was compared to a group of controls taken from a different long-term learning investigation (Papagno & Vallar, 1992). As a result, these controls were unlikely to be matched with SC on variables such as nonverbal ability and age. Despite this limitation, the case of SC provides yet further evidence that the phonological loop "plays a crucial role in the acquisition of novel items" (Trojano & Grossi, 1995; p. 350).

A somewhat different case has been reported in detail by Baddeley (1993). SR was a young graduate student who presented with a developmental deficit of the phonological loop. Although SR's phonological loop system was argued to be qualitatively normal, as evidenced by his sensitivity to the PSE and WLE, his auditory memory span was only four digits, and he was only capable of accurately repeating single syllable nonwords. SR showed normal performance for the learning of word-word pairs, but his performance on word-nonword pairs (nonwords were Finnish words) was dramatically lower than that for a group of matched controls. SR was therefore deemed to have a reduced phonological short-term store capacity which led to an inability to learn the phonological forms of unfamiliar words.

However, there was one striking difference between SR and both PV and SC. PV and SC had acquired their phonological loop impairments after childhood and so it was presumed they acquired their native vocabulary at a rate comparable with typically developing children. In contrast, SR seemed to have a developmental deficit of the phonological loop, suggesting that the process of vocabulary acquisition would have suffered due to this impairment, resulting in lower vocabulary knowledge than would be expected for his age. However, this assumption contradicted his impressive academic achievements. This initially provided a challenge for the PSTM hypothesis. However, the results are reconciled somewhat by SR's sophisticated use of visual coding and mnemonics to compensate for his poor PSTM. Like his controls, SR did attempt to use semantic coding where possible but he was unable to use the strategy of rote rehearsal which the controls frequently reported adopting. Importantly, SR's high general intelligence and motivation were assumed to offset vocabulary acquisition limitations posed by his phonological loop impairment.

Summary

Neuropsychological patients with deficits in the phonological loop have provided convincing support for the PSTM hypothesis (e.g. Baddeley, 1993; Baddeley et al., 1988; Trojano & Grossi, 1995; Vallar & Baddeley, 1984a, 1984b). These patients have shown a consistent pattern of results highlighting a specific impairment in the paired-associate long-term learning of novel phonological material, despite normal long-term learning of pairs of familiar words. This failure to learn word-nonword pairs has been taken to reflect the reliance on temporary phonological storage in the absence of long-term lexical or semantic representations of nonwords. Given their defective phonological storage, these patients subsequently fail to create permanent LTM traces for these nonwords. Patients with developmental impairments of PSTM also emphasise the utility of non-phonological learning strategies, such as visual or semantic coding and mnemonics, coupled with high general intelligence, education and determination, in facilitating cognitive and educational achievements despite limitations in acquiring new phonological forms (e.g. Baddeley, 1993).

1.7.1.2 Evidence from Developmental and Experimental Studies with Typically Developing Children

Although studies with neuropsychological patients have revealed a role for PSTM in the process of vocabulary acquisition, a more direct exploration of this relationship would be to conduct studies with children. Childhood represents an intensive period of vocabulary acquisition and a stage in life when large individual differences in cognitive abilities are generally observed (Gathercole & Adams, 1993). The application of longitudinal studies offers a method with which to assess the strength and direction of this relationship, along with the potential to identify variables which may predict later PSTM and vocabulary knowledge abilities.

A developmental association between PSTM skills and vocabulary acquisition was first established following an influential longitudinal study conducted by Gathercole and Baddeley (1989a). A sample of 104 children were tested upon entering school at 4 years of age, and again one year later, on a number of measures including; PSTM (assessed by nonword repetition), receptive vocabulary, nonverbal ability and reading ability. A significant association between PSTM and vocabulary knowledge was found at both 4 and 5 years of age, even when the more general cognitive factors of age and nonverbal ability were statistically controlled. Furthermore, nonword repetition skills at age 4 were found to significantly predict vocabulary scores at age 5; this finding remained when controlling for age, nonverbal ability and vocabulary scores (at age 4 years). Despite the correlational nature of these findings, Gathercole and Baddeley (1989a) proposed that PSTM skills play a causal role in vocabulary development (see also Gathercole & Baddeley, 1989b). A significant and reliable association between PSTM skills and vocabulary knowledge has also been reported in preschool children (e.g. Gathercole & Adams, 1993). Although this study did not provide evidence of a direct relationship between PSTM and vocabulary knowledge, it does demonstrate that PSTM skills can be reliably tested in 2- and 3 year olds.

In attempt to specify the causal direction of the relationship between PSTM and vocabulary knowledge, Gathercole et al. (1992) extended the results of Gathercole and Baddeley (1989a). A sample of 80 children was assessed at ages 4, 5, 6 and 8 years on the same cohort of measures used in the earlier study. As expected, PSTM and vocabulary knowledge were significantly associated at each of the four age waves, even after partialling out factors of age and nonverbal intelligence. Of more interest were the results of a cross-lagged partial correlational analysis of the data[1]: nonword repetition at age 4 was significantly associated with vocabulary knowledge at age 5; however, vocabulary knowledge at age 4 failed to predict nonword repetition scores one year

---

[1] The technique of cross-lagged correlation (Crano & Mellon, 1978) involves comparing the correlation between two variables across two time points in each direction. For example, the correlation between PSTM skills at age 4 and vocabulary knowledge at age 5 would be compared with the correlation between vocabulary knowledge at age 4 and PSTM skills at age 5. The correlation should be stronger in the causal than in the noncausal direction.

later. Interestingly, however, this pattern of results was reversed for the remaining time intervals (ages 5 and 6 years, and ages 6 and 8 years); the partial correlations between early vocabulary knowledge and later PSTM skills were greater than the converse correlation between early PSTM skills and later vocabulary knowledge.

The results from Gathercole et al. (1992) suggest that the causal nature of the association between PSTM and vocabulary knowledge may change during the course of development. Between the ages of 4 and 5 years, PSTM skills appear to play a predominant role in vocabulary acquisition; however, from 5 to 8 years of age, linguistic knowledge exerts a significant influence on children's PSTM skills. The authors propose that younger children with good PSTM skills "produce phonological memory traces that are highly discriminable and persistent" (p. 896) which, in turn, increases the chances that these memory traces will result in stable and permanent phonological representations, which later become linked with their semantic referents. For children beyond 5 years of age, this reliance on PSTM to acquire new words declines in the face of expanding vocabularies; children are able to use analogies with existing vocabulary items to learn the phonological forms of new words. Indeed, this latter explanation is supported by the observation that linguistic factors, such as the subjective rating of the wordlikeness of nonwords, also independently influence PSTM skills (e.g. Gathercole et al., 1991a; Gathercole, 1995).

Gathercole et al.'s findings (1992) suggest that the earlier claim that PSTM influences vocabulary knowledge in a unidirectional manner is oversimplistic (e.g. Gathercole & Baddeley, 1989a). Instead, this relationship seems more accurately portrayed as a developmental interplay between the two skills. Similar findings have also been reported by Gathercole, Tiffany, Briscoe, Thorn and The Avon Longitudinal Study of Parents and Children (ALSPAC) team (2005) in a sample of 8 year olds.

Furthermore, Gathercole, Service, Hitch, Adams and Martin (1999) report a significant correlation between PSTM skills, indexed by nonword recall and digit span, and vocabulary knowledge in a sample of 13-year olds after controlling for nonverbal ability. The authors argue that there is indeed a developmental continuity in the contribution of PSTM skills to vocabulary development. Nevertheless, this claim would have been further substantiated if conducted across two time points to allow for cross-lagged correlational analysis to determine the causal direction of the relationship (e.g. Gathercole et al., 1992). Interestingly the relationship between PSTM and vocabulary knowledge has also been shown to extend to adults (Atkins & Baddeley, 1998; Gupta, 2003). Atkins & Baddeley (1998) found that the rate of learning visually presented word-nonword pairs (nonwords were Finnish words) was highly associated with PSTM skills, as assessed by nonword repetition, but not to visuospatial STM skills, as assessed by memory for patterns.

Despite the advantages of longitudinal studies, these studies suffer from an unavoidable design limitation in that they are governed by correlational data. This inevitably constrains any interpretation; there is always the possibility that a correlation may actually reflect the influence of a third, unidentified or uncontrolled for, variable. Indeed, Gathercole and Baddeley (1993) point out that the linguistic and social environment of a child could bias the observed developmental association between PSTM and vocabulary acquisition.

In an attempt to control for environmental experience, Gathercole and Baddeley (1990b) conducted a laboratory-based study in which the long-term phonological learning abilities of 5 year old children with high and low PSTM skills were compared. PSTM ability was assessed with a nonword repetition task and the two groups were matched on nonverbal ability. The children had to learn the names of two sets of plastic toy animals. Crucially, one set of toys were given familiar names, such as *Simon* and *Peter*, and the other set of toys were assigned phonologically unfamiliar names, such as *Meton* and *Pimas*. In line with the PSTM hypothesis, the low repetition group took a significantly greater number of trials to learn the unfamiliar names compared with the high repetition group. No difference was found between the two groups in the speed of learning the familiar names of the toys. In addition, when retested for these names 24 hours later, the low repetition group remembered fewer names than the high repetition group.

Gathercole and Baddeley (1990b) concluded that, even when controlling for the amount of exposure to new phonological words, PSTM contributes directly to the long-term learning of unfamiliar material. They postulated that poor PSTM abilities may result in poorly specified phonological representations; these may be influenced by either the insufficient quality of the encoding of phonological memory traces or by experiencing more rapid decay functions of the encoded phonological memory traces.

A similar pattern of findings emerged from a study conducted by Michas and Henry (1994) in which 5 year old children were taught the names and definitions of three novel words, such as *gondola, minstrel* and *platypus*. For each new word, three definitions were also provided that explained what the new word referred to. For example, the definitions provided for the word platypus were '*it has a flat nose*', '*it likes to swim in water*' and '*it eats worms*'. The children were assessed on measures of PSTM (nonword repetition and nonword memory span), spatial memory and vocabulary knowledge. Word learning ability was assessed by measures of word production, word comprehension and recall of the definitions. The results revealed that both PSTM measures were equally significantly correlated with vocabulary knowledge, even after the effects of age and spatial memory were controlled for. Moreover, both PSTM measures were found to be significant predictors of word learning ability when measured by word production, even after controlling for age and spatial memory. Michas and Henry (1994) concluded that PSTM is required to establish long-term phonological representations of new words and that "producing the name for a lexical item is clearly dependent upon having a phonological representation" (p. 160).

A more recent study investigated the PSTM hypothesis using the paired-associate paradigm. Gathercole et al. (1997) tested five year old children on their ability to learn either word-word pairs, such as *table-rabbit*, or word-nonword pairs, such as *fairy-bleximus*, over the course of five trials. Two further tasks involved learning new words (nonwords) in the context of a story in order to provide an even closer analogue to real language learning. In these tasks, the child was told a story which introduced novel words, such as *drattle,* and their meanings, such as *red sticky grass*. The child had to either recall the label for the novel word when given its definition (recall of new word task), or recall the definition of a novel word when given the novel word label (recall of definitions task). Measures of PSTM (digit span and nonword repetition), vocabulary knowledge and nonverbal ability were also obtained.

Both PSTM measures were found to be highly and significantly associated with the rate of learning the word-nonword pairs, but not with the rate at which the word-word pairs were acquired. This pattern of results remained after variance attributable to age, nonverbal ability and vocabulary knowledge were taken into account. Furthermore, both story learning tasks correlated with nonword repetition, although digit span only correlated with the recall of new names task. Gathercole et al. (1997) concluded that PSTM abilities constrain the ease with which new phonological information is acquired, whereas the ability to learn already familiar information proceeds independently of PSTM ability.

Summary

Evidence from longitudinal studies with children have converged upon the view that there is a strong and reliable relationship between PSTM abilities and vocabulary knowledge (e.g. Gathercole & Adams, 1993; Gathercole & Baddeley, 1989a, 1990b; Gathercole et al., 1992, 1997; Gathercole, Service et al., 1999; Michas & Henry, 1994). Moreover, the influence of PSTM skills on vocabulary development remains when controlling for environmental exposure to new words (e.g. Gathercole & Baddeley, 1990b). In addition, the relationship between these two variables has been shown to change during development: PSTM skills appear to influence vocabulary knowledge in early childhood, whereas vocabulary knowledge exerts an influence on PSTM abilities from the age of 5 years onwards (e.g. Gathercole et al. 1992; see also Gathercole, 2006a). Despite this developmental change, PSTM and vocabulary knowledge still remain associated in later childhood (e.g. Gathercole, Service et al., 1999). Furthermore, the relationship between PSTM and new word learning has also been extended to adults (e.g. Atkins & Baddeley, 1998).

1.7.1.3 Evidence from Second Language Learning

Given the evidence that PSTM skills influence vocabulary acquisition in children acquiring their native language (e.g. Gathercole & Baddeley, 1989a, 1990b; Gathercole et al., 1992, 1997; Michas & Henry, 1994), it seems plausible to suggest such a relationship would exist when learning the vocabulary of a second language. A number of studies have been conducted examining the extent to which PSTM predicts later learning of a second language (e.g. Cheung, 1996; Ellis & Beaton, 1993; Masoura & Gathercole, 1999, 2005; Papagno & Vallar, 1995; Service, 1992; Service & Craik, 1993; Service & Kohonen, 1995).

Service (1992) carried out a longitudinal study on a sample of 9- and 10-year old Finnish children who were due to commence learning English through their school curriculum. The children's PSTM skills were assessed prior to starting the course using a nonword repetition task composed of English-sounding nonwords. Children's ability on this task was found to be a very strong predictor of their English grades three years later. On the basis of these findings, Service (1992) concluded that the association between nonword repetition and foreign language learning is mediated by the contribution of PSTM to vocabulary acquisition. Interestingly, these results contrast with the findings from Gathercole et al. (1992) who argued that the contribution of PSTM to vocabulary acquisition was minimal in children over the age of five years. Instead, the findings from Service (1992) suggest that PSTM skills may remain important in later childhood, especially when learning the vocabulary of a new language (Gathercole & Baddeley, 1993).

In a later study, Service and Kohonen (1995) extended Service's (1992) earlier findings by demonstrating that Finnish children's abilities to learn the vocabulary of a foreign language (English) was directly mediated by their nonword repetition skills, even after controlling for general academic achievement. Other aspects of foreign language learning, such as written production of English, were not as strongly associated with nonword repetition skills.

The studies reported by Service (1992) and Service and Kohonen (1995) suggest that learning the vocabulary of a second language is constrained by an individual's PSTM skills (Gathercole & Thorn, 1998). This claim is further attested by the finding that individuals with exceptional language learning abilities have correspondingly superior PSTM skills. Papagno and Vallar (1995) compared a group of polyglots (i.e. individuals who were proficient in at least three languages, including their native language) with a group of non-polyglots. The groups were tested on a number of PSTM skills, assessed by digit span and nonword repetition, and paired-associate learning. The groups performed similarly on measures of general intelligence, nonverbal ability, visuospatial STM and learning pairs of familiar Italian words. However, the polyglots performed significantly better than the non-polyglots on the digit span and nonword repetition tasks, and also on the learning of word-nonword pairs (nonwords were Russian words).

A number of studies have examined the impact of long-term knowledge of the structure of a second language on learning the vocabulary of that language (e.g. Cheung, 1996; Masoura & Gathercole, 1999, 2005; Thorn & Gathercole, 1999, 2001). Cheung (1996) assessed a group of 12-year old Hong Kong children, who had been learning English since early childhood, on their ability to learn the pronunciation and Cantonese translations of three English words (*egregious, succulent* and *jocular*). Measures of nonword span (nonwords conformed to English phonology), word span (English words) and nonverbal ability were also taken. The results showed that nonword span was a unique predictor of the ability to learn the English words. To further investigate whether long-term phonological knowledge of the English language contributed to the children's ability to learn the new English words, the children were split into two groups based on their scores on an English vocabulary test. Children with scores above the median were classified as the high-vocabulary group (high phonological knowledge) and children with scores below the median were classified as the low-vocabulary group (low phonological knowledge). Nonword span was shown to be a significant predictor of the ability to learn the new English words only for the children in the low-vocabulary group. No such association was found for children in the high-vocabulary group. Cheung (1996) concluded that the contribution of PSTM to second language learning declines as long-term phonological knowledge of that language increases. This finding is somewhat reminiscent of Gathercole et al.'s (1992) results with children learning their native language.

In line with Cheung (1996), Masoura and Gathercole (2005) also reported evidence for the influence of existing vocabulary knowledge on second language learning. They assessed paired-associate learning, involving English picture-word pairs, in a sample of Greek children aged between 9 and 11 years who had been studying English as a second language for an average of three years. A significant association between children's PSTM skills and current knowledge of English vocabulary was found, which remained when controlling for age and nonverbal ability. However, more importantly, the children's ability to learn the picture-word pairs, as indexed by the mean number of words correctly recalled and the mean number of trials to learn a new word, was found to be independent of their PSTM skills, but strongly related to the extent of their existing English vocabulary knowledge.

This finding demonstrates that existing long-term phonological knowledge mediates vocabulary acquisition to an increasing extent as an individual's familiarity with the language expands; as a result, there is a reduced reliance on PSTM to aid the learning of new vocabulary. Moreover, a further study by Masoura and Gathercole (1999) found a similar pattern of results. Native and second language vocabulary shared a close relationship which could not be accounted for in terms of PSTM skills.

Further support for this view comes from studies on bilingual individuals (e.g. Thorn & Gathercole, 1999, 2001). Thorn and Gathercole (1999) compared two groups of children: native English-French bilingual children (who had acquired both languages in parallel from birth) and non-native English-French bilingual children (who had started to learn their second language from the age of 2 years). The children were tested on measures of PSTM, as assessed by digit span and nonword repetition, vocabulary knowledge and nonverbal ability. The results for the native bilingual group showed comparable levels of performance on both PSTM measures and on tests of vocabulary knowledge in each language. In contrast, the non-native bilingual group showed a first language superiority for vocabulary knowledge and nonword repetition; that is, children were better at repeating unfamiliar nonwords which conformed to the sound structure of their first language. However, this first language superiority was not found for digit span. The two groups did not differ on nonverbal ability. Thorn and Gathercole (1999) conclude that knowledge of the phonological structure of a language influences PSTM ability, suggesting that "phonological short-term memory is not a language-independent system but, rather, functions in a highly language-specific way" (p. 303). This first language superiority for measures of PSTM has also been shown to extend to adults (e.g. Thorn & Gathercole, 2001).

Summary

Evidence suggests that a strong association exists between PSTM abilities, as assessed by nonword repetition and paired-associate learning tasks, and the acquisition of vocabulary in a second language (e.g. Papagno & Vallar, 1995; Service, 1992; Service & Kohonen, 1995). However, later research has highlighted the influence of long-term language knowledge on second language vocabulary learning (e.g. Cheung, 1996; Masoura and Gathercole, 1999, 2005; Thorn & Gathercole, 1999, 2001). Specifically, children with more expansive second language vocabularies do not require a large contribution from PSTM; in contrast, children with only limited vocabulary knowledge in a second language need to utilise PSTM to a greater extent (e.g. Cheung, 1996; Masoura & Gathercole, 2005). Furthermore, non-native bilingual children and adults show a first language superiority for measures of PSTM ability (e.g. Gathercole & Thorn, 1999, 2001). These findings suggest that the relationship between PSTM and long-term phonological knowledge in second language learning is bi-directional in nature (Masoura & Gathercole, 2005).

1.7.1.4 Evidence from Children with Specific Language Impairment

The majority of the studies reviewed so far have concentrated on gaining support for the PSTM hypothesis with typically-developing children. However, examining the PSTM abilities of children who have developmental language disorders provides a possible further avenue of investigation into the role of PSTM in language learning. One developmental language disorder that has been extensively studied in the PSTM and vocabulary acquisition literature is that of Specific Language Impairment (SLI).

SLI is a developmental language disorder characterised by a deficit in language skills in the absence of any general cognitive impairments, sensory deficits or social deprivation (Archibald & Gathercole, 2006; Gathercole, 2006). Children with SLI have difficulties in various areas of language including processing the rapid transitional information contained within a speech signal (Tallal & Piercy, 1975; Tallal, Stark & Mellitus, 1985), syntax and morphology (Bishop, 1992), and vocabulary development (Stark & Tallal, 1981; see Leonard, 1998, for a review). Research has also shown that children with SLI have marked deficits in PSTM skills, most notably nonword repetition (e.g. Archibald & Gathercole, 2006; Bishop, North & Donlan, 1996; Botting & Conti-Ramsden, 2001; Dollaghan & Campbell, 1998; Gathercole & Baddeley, 1990a; Montgomery, 1995), and the long-term learning of phonological forms of new words (e.g. Dollaghan, 1987; Ellis Weismer & Hesketh, 1996).

Gathercole and Baddeley (1990a) were amongst the first to suggest that the language impairments observed in children with SLI were a result of PSTM deficits. They assessed the performance of a group of SLI children (mean age of 8 years) on a number of PSTM skills. The SLI group were compared to two control groups: a group matched for nonverbal age (mean age of 7 years) and a group matched for language abilities (mean age of 6 years). The SLI group were found to be particularly impaired on a test of nonword repetition compared to both nonverbal and language control groups. However, although the three groups performed at similar levels for nonwords containing up to two syllables, the SLI group performed significantly more poorly than both control groups on three- and four syllable nonwords. The nonword repetition abilities of the SLI group were subsequently estimated to correspond to those of an average 4 year old, representing a four year lag behind their chronological age of 8 years. More recent studies have provided further evidence for the increased sensitivity of the SLI group to increasing length of nonwords (e.g. Archibald & Gathercole, 2006; Bishop et al., 1996; Dollaghan & Campbell, 1998; Montgomery, 1995).

Gathercole and Baddeley (1990a) concluded that the SLI children's language impairments were caused by their severe deficits in the temporary storage of phonological information. Moreover, these deficits were not attributable to impairments in auditory perceptual processing, articulation rates, or a failure to encode material phonologically. In contrast, Van der Lely and Howard (1993) reported no differences in the nonword recall of children with SLI and language-matched controls. However, this finding may be a consequence of using only monosyllabic nonwords. Children with SLI have also been found to have selective problems in learning phonologically novel names of new concepts, compared to age-matched controls; in contrast, these children were unimpaired in learning non-phonological aspects of new words (Dollaghan, 1987).

Interestingly, the severe nonword repetition deficits observed in children with SLI have been reported to have a genetic basis. Bishop et al. (1996) compared the nonword repetition abilities of monozygotic (MZ) and dizygotic (DZ) twins in which at least one child had received a diagnosis of SLI. The results revealed that the SLI group performed poorly on the nonword repetition task. More importantly, nonword repetition scores were found to be significantly lower in the MZ compared to the DZ co-twins. A number of other twin studies have reported similar findings (e.g. Bishop, Adams & Norbury, 2006) leading to the conclusion that nonword repetition deficits in SLI are highly heritable and therefore provide an effective phenotypic marker of the disorder (Gathercole, 2006a). Intriguingly, Bishop et al. (1996) also reported nonword repetition deficits in older children whose language impairments had resolved. This suggests that impairment in nonword repetition is not the causal factor in SLI. However, the authors propose that children are capable of capitalising on their general intellectual abilities to compensate for early language deficits.

Summary

A number of studies have shown that children with SLI have impairments in PSTM skills, most notably nonword repetition. Moreover, the extent of this impairment in nonword repetition increases as the length of nonwords increases (e.g. Archibald & Gathercole, 2006; Gathercole & Baddeley, 1990a; Bishop et al., 1996). Research has also suggested that the nonword repetition deficit observed in children with SLI is highly heritable and represents a phenotypic marker of the disorder (e.g. Bishop et al., 1996, 2006).

1.7.1.5 Evidence from Experimental Word Learning Studies with Adults

Given the accumulating evidence that PSTM is important in acquiring new words, it would be predicted that any interference within PSTM would result in difficulties with the long-term learning of new phonological material (Baddeley et al., 1998). It is well documented that a number variables, such as word length, phonological similarity and articulatory suppression, interfere with the operation of the phonological loop in well specified ways (e.g. Baddeley et al., 1975; Conrad & Hull, 1964; Murray, 1967, 1968; Salamé & Baddeley, 1982). With this in mind, experimental word learning studies conducted with adults have sought to investigate how the learning of novel words is affected by these variables (e.g. Papagno, Valentine & Baddeley, 1991; Papagno & Vallar, 1992).

In a series of experiments, Papagno et al. (1991) examined the effect of articulatory suppression on the learning of familiar and unfamiliar material using the paired-associate paradigm with Italian and English participants. Participants were tested on their ability to learn word-word and word-nonword pairs (nonwords were unfamiliar Russian words) under conditions of articulatory suppression or manual tapping[2] with auditory and visual presentation. If it is the case that PSTM mediates the

---

[2] Manual tapping was used as a non-verbal control to the articulatory suppression condition in an attempt to equate any general processing demands across the two conditions.

long-term learning of unfamiliar material, then engaging in articulatory suppression, which is known to prevent the process of subvocal rehearsal (e.g. Murray. 1967, 1968), would be expected to interfere with the learning of word-nonword pairs, regardless of the modality of presentation

The results with Italian participants were clear (Experiments 1 and 2). Articulatory suppression had little effect on the learning of pairs of familiar Italian words, but had a deleterious effect on the learning of word-nonword pairs with both visual and auditory presentation. However, a different pattern of results was found with English participants. Although articulatory suppression failed to impair the learning of pairs of familiar English words, no selective disruption to word-nonword learning was observed in either modality (Experiments 3 and 4). In an attempt to explain this latter result, a further group of Italian and English participants undertook an assessment of the associative value of the Russian words and Italian-Russian or English-Russian pairs. Participants were required to produce, in their native language, as many words as possible that were suggested by the Russian words or native language-nonword pairs. The results found that a considerably higher percentage of English participants were able to produce an association to the Russian words and English-Russian pairs within the first 5 s of their presentation. It was concluded that English participants were better able to utilise semantic coding of the Russian words to produce meaningful associations between the word-nonword pairs. As a result, two subsequent experiments were conducted with English participants, in which the association values of the nonwords were reduced by using CVC nonwords, such as *jorfap*, or Finnish words. These succeeded in producing the expected selective disruption to the learning of word-nonword pairs (Experiments 6 and 7).

Taken together, these results provide support for the PSTM hypothesis. Under circumstances in which there are no existing lexical or semantic representations available in LTM to facilitate learning, as is the case with unfamiliar material, participants are forced into relying primarily on phonological coding in order to construct permanent representations of this material. In contrast, PSTM does not appear to play much of a role in learning familiar material for which non-phonological learning codes, such as semantic coding, can mediate learning. Indeed. the use of semantic, as opposed to phonological, coding in the long-term learning of familiar material has been previously reported (Baddeley, 1966b; Dale & Baddeley, 1969).

In a similar vein, Papagno and Vallar (1992) examined the influence of phonological similarity and word length on the learning of familiar and unfamiliar material. Given that both these variables are known to interfere with the operation of the phonological loop, they should disrupt the learning of unfamiliar material if this learning is mediated by PSTM. The expected patterns of results were obtained. Neither variable disrupted the learning of word-word pairs, confirming that participants utilise lexical-semantic codes which circumvent reliance on PSTM. However, both phonological similarity and word length slowed down the learning of the word-nonword pairs, providing further evidence that PSTM is relied upon when acquiring new words.

Summary

Experimental word learning studies conducted with adults have shown that variables known to interfere with the operation of the phonological loop, such as articulatory suppression, phonological similarity and word length, impair the learning of word-nonword pairs (e.g. Papagno et al., 1991; Papagno & Vallar, 1992). This has been taken as evidence that PSTM mediates the learning of unfamiliar material. In contrast, the learning of word-word pairs is not affected by such variables, suggesting that the learning of familiar material relies on the use of non-phonological codes, such as semantic codes.

1.7.2 Section Summary

The preceding sections have presented a large body of evidence in support of the PSTM hypothesis. This evidence has converged on the view that the ability to temporarily store unfamiliar phonological representations influences the ability to acquire new word forms. A large proportion of this evidence stems from developmental studies conducted with children learning a native language; these have shown a strong association between PSTM skills, typically assessed by digit span and nonword repetition, and vocabulary knowledge (e.g. Gathercole & Baddeley, 1989a; Gathercole et al., 1992; Gathercole, Service et al., 1999). Moreover, PSTM has been shown to be a good predictor of later vocabulary knowledge (e.g. Gathercole et al., 1992). A similar relationship has also been found in children acquiring a second language (e.g. Service, 1992; Service & Kohonen, 1995). Marked deficits in PSTM skills, specifically nonword repetition, have been argued to underlie the language impairments observed in children with SLI (e.g. Bishop et al., 1996; Gathercole & Baddeley, 1990a). The influence of existing language knowledge has also been shown to play a role in vocabulary acquisition during later childhood, suggesting that the reliance on PSTM to acquire new words declines as children's vocabularies expand (e.g. Gathercole et al., 1992). This finding has also been extended to second language learning (e.g. Cheung, 1996; Masoura & Gathercole, 1999, 2005; Thorn & Gathercole, 1999, 2001) These results demonstrate that PSTM skills and vocabulary knowledge develop in a highly interactive manner.

Experimental word learning studies conducted with children, adults and neuropsychological patients have shown that PSTM abilities relate specifically to the long-term learning of unfamiliar words, but not to the learning of familiar words (e.g. Gathercole & Baddeley, 1990b; Gathercole et al., 1997; Masoura & Gathercole, 2005; Michas & Henry, 1994; Papagno & Vallar, 1995). Furthermore, variables which are known to affect the operation of the phonological loop, such as word length, phonological similarity and articulatory suppression, have shown corresponding effects on the paired-associate learning of word-nonword pairs, but not the learning of word-word pairs (e.g. Papagno et al., 1991; Papagno & Vallar, 1992). Individuals suffering with impairments of the phonological short-term store also show an inability to learn the phonological forms of new words (e.g. Baddeley, 1993; Baddeley et al., 1988; Trojano & Grossi, 1995).

Drawing all this evidence together suggests that the ability to temporarily store the phonological forms of new words is important in order for more stable and permanent representations to be constructed. Indeed, when discussing the primary function of the phonological loop, Baddeley et al. (1998) conclude that it is a "highly flexible language learning system in which ... the phonological loop is available to support the construction of more permanent representations of the phonological structure of new words, but in which established knowledge of the language is used to offset this fragile temporary storage component whenever possible" (p. 170).

1.7.3 Challenges to the Phonological Short-Term Memory Hypothesis

Despite the abundance of evidence in support of the PSTM hypothesis (e.g. Baddeley et al., 1998; Gathercole, 2006a), numerous alternative accounts have been proposed which claim to cast doubt on the specific contribution of PSTM skills to the process of language learning (e.g. Bowey, 1996, 1997, 2001; Chiat, 2001, 2006; de Jong, Seveke & van Veen, 2000; Ellis Weismer and Edwards, 2006; Metsala, 1999; Snowling, Chiat & Hulme, 1991; Van der Lely & Howard, 1993). The majority of these studies argue that the nonword repetition task, which is argued to provide a sensitive measure of the quality of an individual's phonological storage, actually taps a range of phonological abilities, such as auditory processing, phonological processing and speech-motor output processing (Gathercole, 2006a). As such, the predominant challenge to the PSTM hypothesis is that a deficit in nonword repetition does not correspond to a specific deficit in PSTM; in turn, PSTM is not a causal factor influencing the process of vocabulary acquisition.

One of the earliest criticisms of the PSTM hypothesis was proposed by Snowling et al. (1991). They suggest that the nonword repetition task should not be perceived as assessing phonological *memory*, arguing that the repetition of nonwords involves many other complex processes, such as phonological segmentation, assembly of articulatory instructions and speech motor planning. As such, they argue that difficulties with these additional phonological skills are responsible for any delay observed in the long-term learning of phonological representations of words (e.g. Snowling, Goulandris, Bowlby & Howell, 1986; Snowling & Hulme, 1989). Snowling et al. (1991) conclude that vocabulary knowledge directly influences PSTM abilities, stating that "children with good vocabulary knowledge are better able to cope with the processing demands of nonword repetition tasks than are children with poor vocabulary knowledge" (p. 373).

In a rebuttal of this criticism, Gathercole and colleagues (Gathercole, Willis & Baddeley, 1991b) point out that a number of their earlier studies have shown that the complex phonological processing skills advanced by Snowling et al. (1991) do not influence nonword repetition performance (e.g. Gathercole & Baddeley, 1990a, 1990b). Gathercole and Baddeley (1990b) demonstrated that children with high- and low nonword repetition abilities were equally capable of accurately repeating unfamiliar names before commencement of the learning trials. Furthermore, Gathercole and Baddeley (1990a) reported that children with SLI were shown to have poor nonword repetition abilities even when the task did not require spoken output. A twin study conducted by Bishop, Adams and Norbury (2006) also report evidence demonstrating that articulatory duration does not explain the link between SLI and nonword repetition ability.

More recent studies have also addressed these criticisms of the PSTM hypothesis. Gathercole, Service et al. (1999) provided evidence that speech output constraints do not mediate the relationship between PSTM and vocabulary knowledge. They showed that the association between PSTM abilities, as indexed by a nonword recognition task in which no spoken output is required, and vocabulary knowledge, was just as strong as that found between these two variables when PSTM is assessed by nonword repetition and digit span, both of which demand spoken output.

However, despite these findings in defence of the PSTM hypothesis, many researchers continue to support the view held by Snowling et al. (1991) that it is the efficiency of phonological processing skills, rather than phonological storage, which underlies the relationship between nonword repetition and vocabulary knowledge (e.g. Bowey, 1996, 1999, 2001, 2006; de Jong, Seveke & van Veen, 2000; Metsala, 1999). These studies have shown that measures of phonological sensitivity[3] – the ability to detect, analyse and manipulate the sound and syllable units in words (Archibald & Gathercole, 2006; de Jong et al., 2000) – such as rhyme detection, phoneme detection and phoneme deletion, correlate with vocabulary knowledge to a similar degree as PSTM, suggesting that a common phonological processing skill underlies vocabulary development.

Bowey (1996) investigated the extent to which PSTM and phonological sensitivity contributed to vocabulary knowledge in a large sample of preschool children (mean age of 5 years). Phonological sensitivity was measured with rhyme oddity and phoneme identity tasks argued to provide a finer analysis of children's phonological

---

[3] The term phonological sensitivity is also commonly referred to as phonological awareness; indeed, the two terms appear to be used interchangeably within the literature (e.g. Bowey, 1996, 2001; Gathercole, 2006a).

processing skills. The results showed that phonological sensitivity was as strongly associated with receptive vocabulary as was phonological memory, even after age and nonverbal ability were taken into account. Bowey (1996) argued this reflected the contribution of an overall phonological processing ability to vocabulary development, suggesting that PSTM does not mediate long-term phonological learning. However, Gathercole and Baddeley (1997) claim that Bowey's (1996) study has several methodological limitations, arguing that inadequacies in its nonword repetition assessment procedure, coupled with a lack of empirical power and sensitivity, account for the low association found between nonword repetition and vocabulary knowledge.

Bowey's (1996) findings were replicated in a subsequent longitudinal study conducted with children 5 years of age (Bowey, 2001). However, contrary to Bowey's (1996) initial findings, Bowey (2001) demonstrated "the strongest evidence yet observed that nonword repetition predicts subsequent vocabulary development after 5 years of age" (p. 459), thereby offering support for the claim that phonological memory capacity contributes directly to vocabulary knowledge. Bowey (2001) concludes that these findings are most suited to the idea of a reciprocal relationship between nonword repetition and vocabulary development.

In addition, Bowey (2001) also interprets her findings in terms of the lexical restructuring hypothesis (Metsala & Walley, 1998; see also Walley, 1993). This hypothesis suggests that during the early stages of vocabulary acquisition, children initially represent new words in a holistic manner relating to associated acoustic or articulatory routines (e.g. Munson, Edwards & Beckman, 2005), rather than as a sequence of individual phonemes. As a child's vocabulary expands during childhood, there is an increasing need to restructure the lexicon's organisation to allow for more economical phonological representations of lexical items. This restructuring process leads to the creation of more specified sublexical phonological representations in terms of syllables and phonemes (Bowey, 2001; Gathercole, 2006a). According to some authors, it is this lexical restructuring process that is responsible for good performance on nonword repetition tasks (e.g. Metsala, 1999; Metsala & Walley, 1998); hence supporting the claim that vocabulary knowledge influences nonword repetition performance. Support for this view comes from a study conducted by Metsala (1999), who reported a strong association between nonword repetition and vocabulary in

3- to 5 year olds. Moreover, the shared variance of this association was accounted for by phonological awareness measures, as indexed by phoneme blending and initial phoneme isolation tasks, and were not due to PSTM. Interestingly, Metsala (1999) remains open to the proposal that the relationship between rate of vocabulary acquisition and the size of an individual's vocabulary, when measured by paired-associate learning of unfamiliar words, is mediated by PSTM.

Research conducted by de Jong, Seveke and van Veen (2000) provides further support for the idea that additional phonological processing factors, in particular phonological sensitivity, contribute to the acquisition of new words. The authors adopted a slightly modified version of the experimental word learning task originally used by Gathercole and colleagues (Gathercole & Baddeley, 1990b; Gathercole et al., 1997). A group of children (mean age of 5 years) were required to learn the names of unfamiliar cuddly toys. The names were either phonologically unfamiliar, such as *Mobbart* or *Rafin*, or phonologically familiar, such as *Thomas* or *Martin*. They found that phonological sensitivity, as measured by sound categorisation and sound identity tasks, was related to the paired-associate learning of phonologically unfamiliar words, but not to the learning of pairs of familiar words, even after controlling for the factors of age, nonverbal intelligence, vocabulary and letter knowledge. PSTM was also related to the paired-associate learning of the unfamiliar names, although this relationship was eliminated when controlling for nonverbal intelligence and letter knowledge.

In a second experiment, de Jong et al. (2000) examined the effect of phonological sensitivity training on the learning of familiar and unfamiliar words. The authors reported a significant increase in the learning of the unfamiliar words for the group of children receiving training in phonological sensitivity in comparison to a group of control children who received only limited training in sound categorisation. However, the authors do not rule out the possibility that training in phonological sensitivity enhanced PSTM or indeed vocabulary development.

However, in contrast to other reports (Bowey, 1996, 2001; Metsala, 1999), de Jong et al. (2000) failed to find an association between measures of phonological sensitivity and vocabulary, despite finding a close association between PSTM and vocabulary. The authors conclude that "phonological short-term memory might be more strongly associated to current vocabulary, whereas phonological sensitivity is more strongly related to new vocabulary learning" (p. 298).

Despite these challenges, many proponents of the PSTM hypothesis have actively acknowledged the potential contribution of other phonological processing skills to nonword repetition (e.g. Gathercole et al., 1991b; Gathercole, 2006a, 2006b). Indeed, Gathercole (2006a) emphasises that the capacity to store a new phonological representation is not the product of a single factor, arguing that this process is also influenced by "prior factors affecting the initial construction of the phonological representation" (p. 519) and concludes that "multiple perceptual, cognitive, and motor processes constrain both nonword repetition and word learning" (Gathercole, 2006b; p. 610).

However, in a comprehensive review of the challenges to the PSTM hypothesis, Gathercole (2006a) highlights a number of limitations to the view that phonological sensitivity, rather than PSTM, influences vocabulary acquisition. One argument concerns the nature of the tasks assumed to tap phonological sensitivity skills. Although these tasks do assess phonological awareness skills, the majority of these tasks also rely to some extent on phonological storage. Take, for example, the rhyme oddity task in which three words are presented to the child with the requirement to select the odd one out. Successful completion of this task requires the temporary storage of the phonological representations of these words in order to make a phonological comparison between them. Gathercole (2006a) argues that many children will fail this task due to an inability to meet the storage demands imposed by this task, rather than their inability to detect the phonemic aspects of these words. A second limitation for the phonological sensitivity hypothesis concerns its inability to account for the increased sensitivity of children with SLI to increasing length of nonwords (e.g. Gathercole & Baddeley, 1990a). The phonological sensitivity hypothesis offers no explanation as to why a lengthy nonword would require greater phonological sensitivity compared to a shorter (i.e. one- or two syllable) nonword. However, it could be argued that greater demands are placed on the ability to accurately analyse and manipulate the sound units of new words when these new words contain an increasing number of phonemes and/or syllables.

Summary

Despite the accumulating evidence for the role of PSTM in the process of vocabulary acquisition (e.g. Baddeley et al., 1988, 1998; Gathercole & Baddeley, 1990a, 1990b; Gathercole et al., 1992, 1997; Gathercole, Service et al., 1999; Masoura & Gathercole, 1999, 2005; Papagno et al., 1991; Papagno & Vallar, 1992; Service, 1992), a number of alternative interpretations have been forwarded (e.g. Bowey, 1996, 1997, 2001; de Jong et al., 2000; Metsala, 1999; Snowling et al., 1991). The majority of these postulate that nonword repetition encompasses a variety of complex phonological skills, such as phonological segmentation and assembly of articulatory instructions, in addition to PSTM (e.g. Snowling et al., 1991).

A number of studies have reported strong associations between phonological sensitivity skills and vocabulary knowledge, suggesting that PSTM does not mediate the long-term learning of new phonological words (e.g. Bowey, 1996, 2001; Metsala, 1999). These studies have been interpreted with reference to the lexical restructuring hypothesis, suggesting that it is the process of vocabulary growth which accounts for nonword repetition performance (e.g. Bowey, 2001; Metsala, 1999; Metsala & Walley, 1998). Experimental word learning studies have also demonstrated that phonological sensitivity is associated with the paired-associate learning of unfamiliar words to a greater extent than PSTM (e.g. de Jong et al., 2000). However, in defence of these alternative accounts, Gathercole (2006a) highlights a number of methodological limitations with tasks assessing phonological sensitivity. It is suggested that these tasks are dependent on phonological storage and, as a result, any observed influence of phonological sensitivity on vocabulary acquisition inevitably incorporates an element of phonological memory.

1.8 Chapter Summary and Conclusions

A detailed theoretical understanding of PSTM is essential in order to evaluate its role in the learning of new word-forms. With this in mind, the current chapter began by reviewing the concept of the phonological loop component of the WM model (e.g. Baddeley & Hitch, 1974; Baddeley, 1986, 2000a). Evidence was provided to suggest that the phonological loop is capable of accounting for a number of key characteristics of PSTM; these include the effects of word length, phonological similarity, articulatory suppression and irrelevant speech on PSTM performance (e.g. Baddeley, 1986; Baddeley & Salamé, 1986; Baddeley et al. 1975, 1984; Conrad & Hull, 1964; Levy, 1971; Murray, 1967, 1968; Peterson & Johnson, 1971; Salamé & Baddeley, 1983). Although several alternative interpretations have been proposed to explain such effects (e.g. Caplan et al., 1992; Cowan et al., 1997, 2000), and even to discredit the effects themselves (e.g. Jones & Macken, 1993; Jones et al., 2004, 2006), the concept of the phonological loop remains an influential feature of PSTM.

It appears clear from reviewing the PSTM and new word-form learning literature that developmental studies have largely dominated the PSTM hypothesis literature (e.g. Cheung, 1996; Gathercole & Baddeley, 1989, 1990a; Gathercole et al., 1991a, 1992, 1997; Gathercole, Adams et al., 1994; Gathercole, Service et al., 1999; Masoura & Gathercole, 1999, 2005; Michas & Henry, 1994; Service, 1992; Service & Kohonen, 1995; Thorn & Gathercole, 1999). This is understandable considering the relevance of investigating vocabulary acquisition during childhood. These studies have yielded strong evidence of an association between PSTM and vocabulary acquisition, showing that PSTM influences vocabulary knowledge (e.g. Gathercole et al., 1992). Yet, despite their significant contribution to the literature, these studies alone fail to provide a convincing account in support of the PSTM hypothesis given their correlational basis. However, experimental word learning studies, which circumvent this limitation, have provided an additional line of support for the relationship between PSTM and vocabulary acquisition (e.g. Gathercole & Baddeley, 1990b; Gathercole et al., 1997; Michas & Henry, 1994; Papagno & Vallar, 1995)

Whilst a number of studies have investigated new word learning abilities in the form of experimental learning paradigms, mainly the paired-associate learning task, the majority of these have been conducted with children (e.g. Gathercole & Baddeley, 1990b; Gathercole et al., 1997). Only two studies have directly assessed the operation of the phonological loop by manipulating variables known to have well-documented effects on PSTM performance (e.g. Papagno et al., 1991; Papagno & Vallar, 1992), both of which were conducted with adults. Although important insights can be achieved by assessing children's ability to learn unfamiliar phonological words via paired-associate learning tasks, it is more difficult to manipulate PSTM variables when this type of task is given to children. For example, Gathercole et al. (1997) presented 5 year old children with four word-word pairs to learn but only two word-nonword pairs to learn; presenting the children with a larger number of pairs to learn would presumably make the learning task too demanding, thus risking floor effects. Investigating the effects of PSTM variables, such as word length and phonological similarity, would clearly not be appropriate under these circumstances, on the basis that these variables cannot be appropriately implemented when as few as two word-nonword pairs are presented. There are a number of other advantages to conducting experimental word learning studies with adults. For example, adults can generally withstand more intensive investigation than children; as a result, various aspects of experimental design can be implemented. For example; task difficulty can be increased, testing sessions can be extended, tasks can be more complex, and secondary tasks (such as articulatory suppression) can be conducted. These additional design implementations may be particularly useful when examining the detailed mechanisms and processes underpinning the relationship between PSTM and vocabulary acquisition.

This thesis therefore aims to increase our understanding of the role of PSTM in new word-form learning by further investigating the effects of manipulating PSTM variables during a paired-associate learning task conducted with adults. This research is presented in Chapters Three and Four. The second aim of this thesis is to provide a further test of the PSTM hypothesis by investigating whether the Hebb repetition paradigm can be viewed as an experimental analogue to new word-form learning. The Hebb repetition paradigm has so far not been applied to the literature pertaining to PSTM and new word learning. To this end, Chapter Two provides a review of the existing literature on the Hebb repetition paradigm, along with the specific rationale for this research. The experimental research conducted using this paradigm will be presented in Chapters Five and Six.

Chapter Two: The Hebb Repetition Paradigm

2.1 Introduction

Chapter One reviewed a considerable amount of evidence suggesting that PSTM plays an important role in the long-term learning of novel phonological word-forms (e.g. Baddeley et al., 1998). It was also argued in Chapter One that experimental word learning studies provide an additional line of support for the role of PSTM in new word-form learning. Given that a number of these studies utilised the paired-associate paradigm to assess word learning abilities (e.g. Baddeley et al., 1988; Gathercole et al., 1997; Papagno et al., 1991; Papagno & Vallar, 1992), one might ask the question of whether these findings can be extended to other learning paradigms not previously investigated within the context of new word-form learning. If additional paradigms were to succeed in producing results which converge with the existing literature, this would provide further evidence in support of the PSTM hypothesis.

With this in mind, one of the main aims of this thesis is to determine whether an alternative learning method, that of the Hebb repetition paradigm (Hebb, 1961), can be viewed as an experimental analogue of novel phonological word-form learning. Thus, the aim of the current chapter is to provide a review of the existing Hebb repetition literature. The chapter will begin by describing the original Hebb repetition study, before reviewing a number of early studies examining the effect on learning of varying experimental aspects of this paradigm. More recent studies which have utilised the Hebb repetition paradigm to specifically investigate the mechanisms underlying both long-term sequence learning and the learning of serial order will also be reported. A computational model which has attempted to simulate learning in this paradigm will then be briefly discussed. The chapter will conclude by considering whether the Hebb repetition paradigm can be used to examine new phonological word-form learning.

2.2 The Original Hebb Repetition Experiment

In 1961, Hebb designed an experiment with the intention of providing support for his claim that information held in STM was based purely on transient reverberatory activity traces which were presumed to decay rapidly, thus failing to create any permanent structural traces in LTM, particularly when further activities disrupt the consolidation of these transient traces (Hebb, 1949). Hebb's (1961) experiment involved aurally presenting participants with 24 trials of 9-digit sequences in the form of a standard ISR task with one critical manipulation: the same sequence was repeated, surreptitiously, every third trial, with all interleaving trials reflecting unique, or nonrepeated, sequences. Participants were required to immediately recall aloud each sequence in the correct serial order. Contrary to his expectations, Hebb (1961) found that performance on the repeated sequences improved over trials. Moreover, this improvement exceeded performance over trials for the nonrepeated sequences (henceforth referred to as filler sequences). Hebb (1961) attributed any improvement over trials for filler sequences to non-specific task practise effects. On the basis of these findings, Hebb (1961) was forced to conclude that "a single repetition of a set of digits ... produces a structural trace which can be cumulative" (p. 43). This increase in performance over trials for repeated sequences has become known as the Hebb Effect.

2.3 Factors affecting the Hebb Effect

Following Hebb's (1961) novel findings, numerous researchers attempted to replicate, modify and extend this result. The majority of the early studies utilised the Hebb repetition paradigm as a vehicle to assess the interaction between STM and LTM (e.g. Baddeley & Warrington, 1970; Bartz, 1969; Caird, 1964; Cohen & Johnansson, 1967a, 1967b; Cunningham, Healy & Williams, 1984; Heron & Craik, 1964; Kidd & Greenwald, 1988; McKelvie, 1987; Melton, 1963; Sechler & Watkins, 1991; Schwartz & Bryden, 1971). These investigations primarily sought to determine the conditions necessary for the Hebb Effect to emerge. This was achieved by manipulating a range of experimental factors, such as repetition spacing and response requirements, and observing their effects on long-term learning. However, the use of the Hebb repetition paradigm became increasingly scarce following this initial period of research, and by the mid-90's it had disappeared altogether. It was almost a decade later when the paradigm began to experience a revival.

Current research using the Hebb repetition paradigm is beginning to accumulate and many researchers are becoming aware of the potential of this paradigm to shed deeper insights into the complex interface between STM and LTM (e.g. Couture & Tremblay, in press; Cummings, Page & Norris, 2003; Cumming, Page, Norris, NcNeill & Hitch, 2006; Fallon, Dommet & Tehan, 2005; Gagnon, Foster, Turcotte & Jongenelis, 2004; Gagnon, Bedard & Turcotte, 2005; Hitch, Fastame & Flude, 2005; Hitch, Flude & Burgess, in press; Hitch, McNeill, Page, Cumming & Norris, 2006; Page, Cumming, Norris, Hitch & McNeill, 2006; Turcotte, Gagnon & Poirier, 2005). Indeed, when reviewing current research which has utilised the Hebb repetition paradigm, Neath, Brown, Poirer and Fortin (2005) suggest that the paradigm may prove to be "useful in answering a number of key theoretical questions in future" (p. 227). In line with this claim, current research using this paradigm has fallen mainly into two interrelated themes: (i) investigating the mechanisms underlying serial order (e.g. Cumming et al., 2003; Hitch et al., 2005, in press); and (ii) determining the extent to which the Hebb Effect depends on the operation of the phonological loop (e.g. Fallon et al., 2005; Hitch et al., in press; Page et al., 2006).

2.3.1 Effects of Repetition Spacing and Sequence Similarity

One of the earliest investigations following Hebb's (1961) finding was conducted by Melton (1963). Melton (1963) replicated the Hebb Effect using visual presentation of digit sequences with written recall. Moreover, he extended Hebb's (1961) original finding by examining the effect of repetition spacing on the Hebb Effect. This was achieved by manipulating the number of filler sequences intervening between the repeated sequences. A Hebb Effect was found with repetition spacings of two, three and five filler sequences, but not with a repetition spacing of eight filler sequences. Melton (1963) concluded that the Hebb Effect decreases as a function of the number of intervening filler sequences, with the effect disappearing with repetition spacings of more than five filler sequences. Furthermore, he attributed this finding to the incremental build up of retroactive interference between the repeated sequences, such that the larger the number of filler sequences, the more information is presented which subsequently interferes with the formation of a permanent structural trace of the repeated sequence.

If one considers Melton's (1963) findings that cumulative learning of a repeated sequence does not occur when the repeated sequence is interspersed by more than five filler sequences, it is difficult to imagine how this would lend support to the proposal that the Hebb repetition paradigm may be an appropriate investigative tool with which to assess the long-term learning of novel word-forms. The process of vocabulary acquisition would suffer immensely if every new word perceived had to be repeated within close temporal proximity to each other in order for that word to be reliably learned (Cumming et al., 2006). However, Cumming et al. (2006) suggest that Melton's (1963) findings may actually be a consequence of using a closed set of materials given that both repeated and filler sequences contained the digits 1 to 9. This high degree of item overlap between these two types of sequences may have been the catalyst for the interference observed in Melton's (1963) study when repeated sequences were presented too far apart. With this in mind, Cumming et al. (2006) manipulated repetition spacing and the degree of sequence similarity between repeated and filler sequences. Participants were visually presented with 7-item word sequences. In one condition the same set of words were used for both the filler and repeated sequences (overlapping condition); in a second condition, a different set of words were used for the repeated and filler sequences (non-overlapping condition). A further manipulation involved the frequency of sequence repetition: the repeated sequence was presented every third or sixth sequence (i.e. two vs. five intervening trials).

Cumming et al. (2006) found reliable learning of the repeated sequence in the non-overlapping condition, but not in the overlapping condition. No differences were found between the two repetition spacing conditions in either the overlapping or non-overlapping conditions. The authors concluded that the critical factor in generating reliable Hebb Effects is the degree of item overlap between items in the repeated and filler sequences. Moreover, equivalent levels of recall performance were observed for the filler sequences in the overlapping and non-overlapping conditions, suggesting that interference was confined to learning only; the lack of reliable learning in the former condition was therefore not attributable to a build up of proactive interference. In a further experiment, the authors extend their findings to repetition spacings of up to 12 trials in the non-overlapping condition[4]; that is, a reliable Hebb Effect was observed with 11 intervening filler sequences. This further confirmed that the rate of learning a repeated sequence is independent of repetition spacing. Importantly, this latter finding

---

[4] Given that Cumming et al.'s (2006) earlier experiments failed to produce reliable Hebb Effects in the overlapping condition, this condition was not conducted in this later experiment.

appears to support the natural process of vocabulary learning; a novel word does not need to be repeated every few words in order for it to be successfully acquired.

2.3.2 Effect of a Distractor Activity

In response to Melton's (1963) study, Bartz (1969) examined the effect of an interpolated activity on the Hebb Effect. On the basis of Melton's (1963) conclusion regarding retroactive interference, Bartz (1969) predicted that the rate of learning a repeated sequence would decrease if an irrelevant activity was interpolated between the presentation and recall of that sequence. Bartz (1969) manipulated the retention interval between presentation and recall of aurally presented 9-digit sequences (immediate recall, 15 s delayed recall, 30 s delayed recall) and the presence of a distractor activity between presentation and recall (silence vs. shadowing sequences of consonants). Spoken recall was required. The repeated sequence was presented every third trial, as in Hebb's (1961) original experiment. A reliable Hebb Effect was observed for the repeated sequences across all groups, with the largest effect occurring for the delayed recall groups. Comparable rates of learning the repeated sequence were observed for the silent and filled conditions, suggesting that an interpolated activity does not have an effect on the rate of learning a repeated sequence. Performance in the immediate recall group improved up to the fourth repetition, but then ceased improving after that point. Performance on the filler sequences improved over trials, although this was attributed to non-specific practise effects. Surprisingly, Bartz (1969) analysed the repeated and filler sequences separately, therefore making any direct comparison with Melton's (1963) results somewhat limited. A stronger test of the Hebb Effect would have been achieved if these two sequence types were incorporated within the same statistical analysis. Bartz (1969) suggested the learning mechanism involved a search process, whereby current sequences are matched against memory for previous sequences, with a successful match resulting in the strengthening of that sequence.

2.3.3 Effects of Recall and Rehearsal

A number of early studies have examined the effects of recall and rehearsal on the Hebb Effect (Cohen & Johansson, 1967a; 1967b; Cunningham et al, 1984, Kidd & Greenwald, 1988). A well cited study conducted by Cohen and Johansson (1967a) investigated whether an overt recall response was required to produce the Hebb Effect. Participants were aurally presented with 9-digit sequences. During a pre-test phase, the instructions differed across each of three conditions: (A) participants were required to simply listen to each sequence and indicate when they felt their attention wandering, thus no overt recall response was required; (B) participants were required to overtly recall only those sequences which were immediately followed by a signal – importantly, participants were never required to make a spoken response to repeated sequences as these were never followed by a signal; and (C) participants were required to provide a spoken recall response to all presented sequences. In the test phase, a spoken recall response was required to all sequences. The repeated sequence in the pre-test phase was re-presented in the test phase. A Hebb Effect was found only in condition (C) in which participants made an overt spoken response to the repeated sequences.

Cohen and Johansson (1967a) concluded that the 'perception' of a repeated sequence does not leave a sufficiently strong trace to facilitate later recall of that same sequence; only instances in which an overt recall response is made to the repeated sequence will that sequence be learned. However, there is a potential problem with this conclusion: the failure to obtain a Hebb Effect in condition (B) may be confounded by the very nature of the instructions provided. If an overt recall response was not required following presentation of the repeated sequence only, and this repeated sequence is presented every third trial, then it is not unrealistic to suppose that participants may become aware of this response schedule. As a result, participants may feel the repeated sequence can be ignored since it will not require an overt recall response. However, when discussing their findings, the authors state that the majority of participants in condition (B) reported actively attempting to group the digits in each sequence; that this failed to facilitate performance on the repeated sequence was argued to provide further evidence that the perception of a sequence alone is not sufficient to strengthen a specific memory trace.

In a further study, Cohen and Johansson (1967b) examined the role of rehearsal strategies and overt recall on the production of the Hebb Effect. A series of 9-digit sequences were presented auditorily, with the repeated sequence presented every third trial. Participants were instructed to use one of two rehearsal strategies: rehearsing the sequence as three groups of three digits (3-3-3) or as a group of five digits followed by a group of four digits (5-4). Three response conditions were tested: (A) silent rehearsal during sequence presentation followed by a verbal recall response only to those sequences which were immediately followed by a signal (these signalled sequences were always filler sequences, as in Cohen & Johansson, 1967a); (B) silent rehearsal during sequence presentation, followed by a written response to all sequences relating to an estimation of the number of digits remembered; and (C) silent rehearsal during sequence presentation, followed by a verbal recall response to all sequences. Both rehearsal strategies were tested in conditions (A) and (B), whilst only the (5-4) strategy was tested in condition (C). In a further condition (D), a verbal recall response was required following each sequence but with no rehearsal strategy imposed; in this condition the presentation rate was increased from 1 digit/s to 4 digits/s, with the aim of preventing rehearsal during presentation.

Evidence of a Hebb Effect was found in conditions (A) and (B) but only when the (3-3-3) rehearsal strategy was adopted; this suggests that an overt recall response is not required to produce a Hebb Effect, as long as the sequence to-be-learned is rehearsed in groups of three. A Hebb Effect also emerged in condition (C), confirming that, when an overt recall response is made, a rehearsal strategy such as (5-4) leads to cumulative learning of a repeated sequence. Condition (D) failed to produce a Hebb Effect; this was taken as evidence that preventing rehearsal by increasing presentation rate eliminates cumulative learning. Cohen and Johansson (1967b) concluded that rehearsal is the key to transferring information from STM into LTM. Moreover, the authors suggest that rehearsal plus an overt recall response produces stronger structural traces in LTM compared to rehearsal alone.

That a Hebb Effect emerged in condition (A) when rehearsal was in groups of three appears to contradict the findings from Cohen and Johansson (1967a), in which a Hebb Effect failed to emerge under similar experimental conditions (condition B). The only difference between these conditions in the two studies refers to the explicit instruction to rehearse in groups of three in Cohen and Johansson (1967b); no such rehearsal instruction occurred in their earlier study (Cohen & Johansson, 1967a). Unfortunately, Cohen and Johansson (1967b) fail to acknowledge this contrasting result; this is especially surprising given that Cohen and Johansson (1967a) explicitly state that participants attempted to group the digits in condition (B). The production of a Hebb Effect in condition (A) in Cohen and Johansson (1967b) suggests that even if participants are aware of the response schedule (i.e. that an overt recall response is not required following every third trial), as mentioned previously, they still engage in rehearsal during presentation of the repeated sequence. As a result, this leads to cumulative learning of the repeated sequence, even in the absence of an overt recall response.

The results of Cohen and Johansson (1967b) suggested that an overt recall response is not necessarily essential in producing the Hebb Effect, given the use of an effective rehearsal strategy. Cunningham et al. (1984) extended this view by proposing that it is not the requirement for an overt response per se, but that the "active processes of rehearsal and coding" (p. 576) are the critical factors required for long-term retention. Participants were visually presented with pairs of 4-letter (consonants) sequences (e.g. BHFK-QRLM). During presentation, participants were required to name each letter aloud, this was followed by a digit-shadowing retention interval (0, 4, 8 or 12 digits) and finally a recall cue indicating which of the two segments should be recalled. Recall responses were written. Identical pairs of 4-letter sequences were presented on each trial for the repeated sequences, with different pairs presented on each trial for filler sequences. Performance in two conditions was tested. In the same segment condition, participants were cued to recall the same segment on each of four presentations of the repeated sequence (e.g. BHFK). In the switched segment condition, participants were cued to recall the same segment on the first three presentations of the repeated sequence (e.g. BHFK) but were cued to recall the other segment on the fourth presentation of the repeated sequence (e.g. QRLM). Improvement in recall performance was found in the same segment condition, but not in the switched segment condition.

Cunningham et al. (1984) suggested that repeating a sequence during both presentation and recall leads to increased recall for that sequence compared to when the sequence is repeated during presentation alone. The authors concluded that active rehearsal was taking place in the same segment condition as the same segment was tested for recall at each presentation. In the switched segment condition, only passive rehearsal was occurring for the first three presentations of the segment that was eventually cued for recall on the fourth presentation. To this end, Cunningham et al.

(1984) argue that "it is the active rehearsal activity that naturally accompanies *recall responses* that is critical in obtaining the repetition effect (p. 589, italics added).

Kidd and Greenwald (1988) also propose that the Hebb Effect arises as a result of rehearsal. Furthermore, they extend their view to circumstances in which only partial recall of a sequence is required. Participants were presented with a probe recall task in which they were required to rehearse an entire sequence but recall only part of that sequence. The task involved the presentation of 9-digit sequences which were each immediately followed by the presentation of a single (probe) digit from the just presented sequence. Participants were required to recall either (i) the digit immediately preceding the probe digit; or (ii) the digit preceding the probe digit by three serial positions. All recall responses were made via a numeric keyboard. The authors found evidence for a Hebb Effect using this task and so concluded that the beneficial effect of a rehearsal opportunity between presentation and recall of a sequence "does not depend on overt reproduction of a *full sequence*" (p. 274, italics added).

Summary

Studies examining the role of rehearsal and overt recall of a repeated sequence on the Hebb Effect have produced rather mixed results. Earlier studies advocated the importance of an overt response to a repeated sequence, suggesting that the mere perception of a sequence alone is not sufficient to create an enduring memory trace (e.g. Cohen & Johansson, 1967a). Subsequent studies revealed the importance of rehearsal in determining the Hebb Effect (Cohen & Johansson, 1967b). However, these studies differed with respect to the type and amount of rehearsal required to produce the Hebb Effect, with some proposing that an active process of rehearsal was required during both presentation and recall for the Hebb Effect to emerge (Cunningham et al., 1984), whilst others emphasised the importance of rehearsal during presentation only (e.g. Kidd & Greenwald, 1988).

The role of rehearsal in producing reliable long-term sequence learning may pose a challenge for the hypothesis that the Hebb repetition paradigm may represent an analogue of novel phonological word-form learning. Research has shown that children under 4 years of age do not utilise the subvocal rehearsal mechanism (e.g. Flavell et al., 1966; Gathercole & Hitch, 1993; Gathercole, Adams et al., 1994; see Chapter One).

However, by the age of 5 years a child has acquired a vocabulary exceeding 2000 words (e.g. Smith, 1926, cited in Baddeley et al., 1926). This suggests that young children are capable of learning new words despite not being able to rehearse. Thus, if it is the case that the Hebb Effect relies on rehearsal, this suggests that young children would not show any evidence of the Hebb Effect. In turn, this would indicate that the Hebb repetition paradigm may not be a suitable analogue of new word-form learning.

2.3.4 Effect of Sequence Repetition Awareness

The role of sequence repetition awareness in the Hebb Effect is an important variable to examine when considering its relevance to the process of vocabulary acquisition undertaken by children. It can be argued that vocabulary acquisition proceeds, to some extent at least, at an implicit level; children can learn new words without explicit instruction to do so. Furthermore, children are capable of learning the phonological structure of a new word without specific knowledge of its referent (e.g. Gathercole & Baddeley, 1990b; Gathercole et al., 1997). With this in mind, if the Hebb Effect is reliant on participants' awareness of sequence repetition, then the Hebb repetition paradigm may not accurately assess the ability to learn new words. A handful of studies have assessed the impact of repetition awareness on the Hebb Effect (e.g. Hebb, 1961; Hitch et al., in press; McKelvie, 1987; Sechler & Watkins, 1991).

The first systematic investigation of the possible influence of repetition awareness on the production of the Hebb Effect was conducted by McKelvie (1987). The design followed that of Hebb (1961) but with the addition of a post-experimental questionnaire developed with the specific aim of classifying participants' degree of awareness. After completing the experiment, participants were asked if they had noticed anything particular about the experimental procedure, and whether they had noticed repetition of any sequence of digits. Based upon their answer to this latter question, participants were then asked at what point in the experiment they had become aware of repetition, how many sequences they felt were repeated, and how often these were repeated. On the basis of participants' responses to these questions, 19% were classified as 'unaware' and 81% were classified as 'aware'. A median split based on the trial at which onset of repetition awareness was reported was used to further categorise 'aware' participants as 'early aware' (trials earlier than the median), 'middle aware' (trials at the median), and 'late aware' (trials later than the median).

A Hebb Effect was observed for all four groups; moreover, performance on the repeated sequences improved at a similar rate across each of the groups. McKelvie (1987) concluded that the size of the Hebb Effect is independent of the degree of awareness; awareness of sequence repetition does not promote faster learning of a repeated sequence. However, McKelvie (1987) is rather tentative with this conclusion: he acknowledges a number of limitations with his experiment, including the fact that all participants received the same repeated sequence and the same set of filler sequences. His evidence for the Hebb Effect could therefore be a result of something specific about the repeated sequence; a stronger test of the influence of sequence repetition would have been to rotate sequences across participants. Finally, the author questions the validity of his criteria in delineating his awareness groups; alternative criteria could have been undertaken to support his findings. Indeed, one could question the validity of the post-experimental questionnaire itself: the answers obtained are extremely subjective; a more stringent test of awareness may require more objective or covert measures.

A later study by Sechler and Watkins (1991) looked for a functional relationship between the Hebb Effect and repetition awareness. The procedure adopted was rather complex: participants were visually presented with either digit sequences or word sequences and over the course of the experiment were required to (1) provide a written recall response to each presented sequence; (2) provide a recognition judgement after six presentations of the repeated sequence – this involved rating on a 6-point scale whether each of five sequences (the repeated sequence, two filler sequences, and two new sequences) had occurred earlier in the experiment; (3) provide a frequency estimation after all sequences had been presented – this involved estimating the number of times each sequence had been presented; and (4) respond to a questionnaire where the critical question asked participants to indicate which of three statements best described the structure of the experiment: "all sequences presented once", "all sequences presented more than once" or "some sequences presented once and some more than once".

Sechler and Watkins (1991) found evidence for a reliable Hebb Effect for both digit and word sequences. In terms of repetition awareness, all three of their measures provided evidence for awareness of the repeated sequence. Moreover, a comparison of effect sizes revealed larger effect sizes for the effect of sequence repetition on both recognition and frequency estimation tests than on the act of reproducing the sequence (i.e. providing a written response). They therefore suggest that these two awareness measures were more sensitive to repetition than the actual recall of the repeated sequence. As such, the authors appear to define the Hebb Effect in terms of repetition awareness, concluding that their results "failed to provide evidence for the Hebb Effect in the absence of repetition awareness" (p. 389). However, despite making this claim, the authors do acknowledge that their conclusions do not extend to those participants who failed to notice sequence repetition. Indeed, it seems the authors' conclusion is not truly upheld by their results: their conclusion implies that the Hebb Effect does not arise without awareness of repetition. Whilst it is clear from their data that the majority of participants did express an awareness of sequence repetition, they do not specifically test for the presence of a Hebb Effect in those participants who denied awareness of repetition. Furthermore, the authors fail to provide information as to the consistency of participants responses; that is, do participants' responses on the three awareness measures converge, or do they only acknowledge awareness on just one of these measures. Surprisingly, Sechler and Watkins (1991) make no reference to the study conducted by McKelvie (1987).

The studies by McKelvie (1987) and Sechler and Watkins (1991) highlight the difficulties in assessing 'true' levels of repetition awareness. McKelvie's (1987) assessment criteria may be considered rather arbitrary and would benefit from replication and a tighter controlled design. Sechler and Watkins (1991) do attempt to assess repetition awareness using more objective measures, but yet fail to accurately support their conclusions with empirical data. Indeed, the very nature of any form of recognition test or questionnaire may serve to alert participants to sequence repetition retrospectively, thus potentially overestimating their true level of repetition awareness. Such studies serve to highlight the more general problem of probing awareness of repetition after a period of learning.

A more recent study conducted by Hitch et al. (in press) provides tentative evidence that degree of awareness and sequence repetition are not strongly associated. They observed similar patterns of results regardless of whether participants noticed sequence repetition, reflecting the findings of McKelvie (1987). However, Hitch et al.'s (in press) procedure for assessing repetition awareness followed the format of earlier studies (e.g. McKelvie, 1987; Sechler & Watkins, 1991) as participants were asked an open question regarding their awareness of sequence repetition (G. J. Hitch, personal communication, 2007). It may be that Hitch et al.'s (in press) study also suffers from the same confounds as previous investigations (e.g. McKelvie, 1987; Sechler & Watkins, 1991) in terms of whether a post-experimental questionnaire represents a valid measure of repetition awareness. However, the authors offer a further potential insight regarding this variable: participants reported partial awareness of sequence repetition limited to the beginnings and ends of digit sequences. Interestingly, Hitch et al. (in press) also report that repetition awareness decreased under conditions of articulatory suppression and link this finding to the claim that Hebb repetition learning may be considered a form of implicit serial learning (e.g. Gagnon et al., 2004; Seger, 1994; Stadler, 1993; see also section 2.4).

Summary

A number of studies suggest that the Hebb Effect emerges independently of awareness of sequence repetition (e.g. Hitch et al., in press; McKelvie, 1987); that is, the Hebb Effect does not appear to be confined to those participants noticing sequence repetition. However, these studies measure repetition awareness by means of post-experimental questionnaires; this retrospective method of assessment is highly subjective and may produce biased results. A further study claims to have developed more sensitive tests of repetition awareness which are less explicit than the questionnaire method; however, this study fails to provide empirical evidence that the Hebb Effect is dependent upon repetition awareness (e.g. Sechler & Watkins, 1991). The effect of awareness on sequence learning clearly remains an area for further investigation; this poses a challenge considering the overlap between the explicit and implicit learning literature (e.g. Gagnon et al., 2004; Seger, 1994) and the inherent circular problems associated with assessing awareness (e.g. Destrebecqz & Cleeremans, 2002, Gagnon et al., 2005; Sechler & Watkins, 1991). The lack of an effect of repetition awareness on sequence learning may lend initial support, albeit tentative, for the validity of the Hebb repetition paradigm in simulating new word learning in the laboratory.

2.3.5 Effect of Changing Sequence Characteristics

A number of studies have examined the effect of changing aspects of the repeated sequence (e.g. Bower & Winzenz, 1969; Hitch et al., 2005, in press; Schwartz & Bryden, 1971). Bower and Winzenz (1969) investigated the effect of changing the temporal grouping structure of a repeated 12-digit sequence: the temporal grouping structure either remained identical (RS condition) or changed (RD condition) at each repeated presentation. Temporal grouping structure was manipulated by the experimenter reading out each digit name and inserting pauses to define each group (e.g. 16-358 was read as one, six.....three, five, eight; Experiment 3). Written recall responses were required. A significant improvement in recall of the repeated sequence was found for the RS condition only; performance in the RD condition was equivalent to that of the filler sequences and showed no improvement over trials.

The authors explained their results in terms of a ‘reallocation hypothesis’, proposing that the memory trace created following presentation of a sequence is transferred to a ‘location in memory’. If two sequences are coded in similar ways, as in the RS condition, these two sequences will be stored in the same location, resulting in strengthening of that particular memory trace. However, if two sequences are coded in substantially different ways at each presentation, as in the RD condition, these will be allocated different locations in memory, despite maintaining the same underlying sequence; this results in a memory trace equivalent to a once-presented filler sequence.

In a similar vein, Schwartz and Bryden (1971) changed the beginning or the end of a repeated sequence to determine whether this disrupted sequence learning. Following Hebb’s (1961) procedure, five conditions were tested in a training phase with varying manipulations of the repeated sequence: the entire sequence was repeated on each presentation (1); one or two digits at the beginning of the repeated sequence changed at each presentation (2) and (3); and one or two digits at the end of the repeated sequence changed at each presentation (4) and (5). A further condition in which the repeated sequence was never presented acted as a control condition. In the test phase, the entire repeated sequence was presented on five occasions, each separated by two filler sequences. Analyses based on participants' recall responses to the repeated sequence in the test phase showed evidence of Hebb Effects in all conditions[5], except condition (3) where two digits had been changed at the beginning of the repeated sequence. Performance on the repeated sequence in this condition equalled that for the control condition in which the repeated sequence was never presented. Therefore, changing two digits at the beginning of the repeated sequence prevented learning of that sequence.

Schwartz and Bryden (1971) postulate that items, or chunks of items, within the repeated sequence may be 'tagged'. The memory traces created on subsequent presentations of the same repeated sequence are tagged in the same way and consequently interact with memory traces from previous presentations, resulting in an improvement over repeated presentations. If, however, items at the beginning of subsequent repeated presentations are changed, these later sequences will acquire different tags at the beginning of a sequence, thus preventing any interaction with previous memory traces. This account fits well with the reallocation hypothesis forwarded by Bower and Winzenz (1969).

More recently, the effect of changing characteristics of the repeated sequence has been re-examined by several researchers (e.g. Hitch et al., 2005, in press). Hitch et al. (in press) investigated the effect of changing temporal grouping structure on the Hebb Effect following the procedure adopted by Bower and Winzenz (1969; Experiment 3). The results replicated Bower and Winzenz (1969) in that sequence learning was observed in the RS condition. However, in direct contrast to Bower and Winzenz (1969), significant learning was also observed in the RD condition. The authors speculate that potential design differences may be the reason for this discrepancy, suggesting that the temporal groups assigned to the RD condition in their study could have been more similar across repeated presentations, thus leading to sequence learning. Hitch et al. (in press) interpret their results in terms of a computational model of long-term sequence learning (Burgess & Hitch, 1999, 2006) which predicts that

---

[5] The control condition was excluded from these analyses.

temporal information is encoded when learning a sequence (see section 2.5 for an overview of this model and its predictions).

Hitch et al (2005) conducted a complex series of experiments which compared the predictions generated by a number of computational models, each concerned with how serial order is learned. The authors' theoretical motivations for undertaking the study, along with the detailed mechanisms and differing approaches underpinning each of these models of serial order, is outside the scope of this thesis and so will not be discussed in detail here. A brief overview of the more relevant models with regard to the current thesis is discussed elsewhere (see section 2.5). The study is mentioned here with reference to Schwartz and Bryden's (1971) earlier finding that the beginning of a sequence is crucial for the later learning of that sequence. Using a modified version of Hebb's (1961) procedure, Hitch et al. (2005) manipulated the repetition of different fragments of sequence information. They found that significant learning of a repeated sequence (consisting of differing numbers of consonants) was only observed when that fragment of repetition started from the beginning of a sequence; if the repeated fragment of a sequence occurred at the end of a sequence, only minimal learning was observed. This suggests that the "repetition of the first or first and second item in a sequence may be critical for learning items at later positions" (p. 256).

Summary

Studies have shown that the long-term learning of a sequence involves the strengthening of memory traces. Importantly, these memory traces need to be identical in their temporal structure in order to become strengthened (e.g. Bower & Winzenz, 1969; Schwartz & Bryden, 1971), although one study has showed some evidence of long-term learning when the temporal structure of a sequence changes over presentations (e.g. Hitch et al., in press). Several studies have highlighted the importance of the beginning of a sequence; learning appears not to proceed unless the initial portion of a sequence is repeated (e.g. Hitch et al, 2005; Schwartz & Bryden, 1971). These findings arguably relate to the process of vocabulary acquisition; a child may find it difficult to construct a permanent representation of a new word if the precise phonological structure of that word varies over repeated exposures.

2.3.6 Effect of Memory Impairment

The effect of memory impairment on the Hebb Effect may provide a further line of investigation into the role of awareness on sequence learning and the extent to which learning via the Hebb repetition paradigm is implicit in nature. One of the earliest investigations of the effect of memory impairments on long-term learning was conducted by Baddeley and Warrington (1970). Given the then speculative view that amnesic patients have a normally functioning STM, but impaired LTM (e.g. Drachman & Arbit, 1966), the authors investigated the effect of sequence repetition with a group of six amnesic patients whose performance was compared to that of a group of control patients matched on age, intelligence and occupation (Experiment 6). Based on the assumption that performance for the repeated sequence reflects LTM, they predicted that their amnesic patients would fail to show the beneficial effect of sequence repetition. The two groups were auditorily presented with 8-digit sequences for immediate verbal recall, with the repeated sequence occurring every second trial for eight presentations, this was then followed by four consecutive presentations of the repeated sequence.

The results obtained were not as predicted: both amnesic and control patients showed evidence of a Hebb Effect, although the effect in both cases was not particularly strong. Baddeley and Warrington (1970) concluded that their results suggest that either LTM is not involved in the Hebb Effect or that their amnesic patients were "unimpaired on at least one type of LTM" (p. 186). Indeed, McKelvie (1987) proposes that the results of his study suggest that "only *explicit* long-term memory need not be involved" (p. 85; italics added). However, Milberg, Alexander, Charness, McGlinchey-Berrorth and Barrett (1988, cited in Seger, 1994) failed to find evidence for cumulative learning in their amnesic patients when the digit sequences presented were longer than their assessed digit spans. A similar result was obtained in a group of 'memory disordered' patients (e.g. Caird, 1964); however, this particular finding cannot be generalised to all patients with memory impairments as no information was provided about the specific memory problems experienced by these patients.

More recently, Gagnon et al. (2004) assessed sequence learning in a densely amnesic patient (SJ) who had sustained a circumscribed bilateral lesion to the hippocampus. SJ and 12 age-matched controls were given the Hebb repetition task using a variety of materials: digits, words and pseudowords. A novel aspect of their design concerned sequence length; each participant was presented with sequences at their span plus one item. This ensured all participants were presented with sequences exceeding their individual capacities and were in this sense matched on STM ability. SJ was also presented with a serial reaction time (SRT) task. This task is used heavily in the implicit learning literature (e.g. Seger, 1994; Stadler, 1993) as it does not encompass an explicit recall component. In this task, participants have to respond to specific targets located in different positions on a computer screen with a key press corresponding to the target's location, with sequences of locations being repeated. Learning is demonstrated by a faster decrease in reaction time for the repeated sequence compared with filler sequences. Gagnon et al. (2004) found that SJ showed reliable and comparable Hebb Effects across each of the three types of materials. Moreover, SJ's learning rates were comparable with the age-matched controls. SJ also demonstrated significant learning of the repeated sequence in the SRT task.

Given these results, Gagnon et al. (2004) concluded that the "hippocampus has a limited role with respect to implicit learning of recurrent sequences" (p. 877) and that "processes underlying implicit learning are spared in amnesic individuals who have sustained selective hippocampal damage" (p. 877). The authors consider the Hebb repetition task an implicit learning task and, in line with McKelvie (1987), speculate that this learning paradigm does not rely on or require explicit LTM. This latter claim was also supported by SJ's inability to recall any of the repeated sequences following the experiment.

Summary

Several well designed studies have shown that patients with LTM impairments are capable of showing cumulative learning of a repeated sequence (e.g. Baddeley & Warrington, 1970; Gagnon et al., 2004). This suggests that the learning observed in the Hebb repetition paradigm can proceed at an implicit level and without explicit awareness of sequence repetition. Indeed, it has been argued that sequence learning does not rely on explicit LTM (e.g. Baddeley & Warrington, 1970; Gagnon et al., 2004; McKelvie, 1987). More recent evidence suggests that the hippocampus does not play a role in the implicit sequence learning observed in patients with selective hippocampal damage (e.g. Gagnon et al., 2004).

2.3.7 Effects of Articulatory Suppression and Phonological Similarity

Several studies have examined the effects of articulatory suppression and phonological similarity on the Hebb Effect (e.g. Fallon et al., 2005; Hitch et al., in press; Page et al., 2006). These are important manipulations to investigate if the Hebb Effect is assumed to depend on the phonological loop, thereby permitting the Hebb repetition paradigm to be used as a potential laboratory analogue to assess the long-term learning of novel phonological word-forms. As described in Chapter One, the concept of the phonological loop is capable of explaining the effects of numerous variables on PSTM, primarily word length, phonological similarity and articulatory suppression. Moreover, the phonological loop successfully deals with the presentation modality effects of articulatory suppression on the WLE and the PSE. If long-term sequence learning via the Hebb repetition paradigm is dependent upon the operation of the phonological loop, then the same pattern of effects should be observed using this paradigm. Indeed, these findings would be predicted given the results of earlier studies which suggest that the Hebb Effect depends on some form of rehearsal (e.g. Cohen & Johansson, 1967b; Cunningham et al., 1984, Kidd & Greenwald, 1988). However, a limitation of these earlier studies lies in their technique for preventing rehearsal, namely increasing item presentation rate. It is now known that articulatory suppression is a more effective method of preventing subvocal rehearsal, particularly when articulatory suppression is required throughout presentation and recall (e.g. Baddeley et al., 1984; Murray, 1967).

A series of experiments have been conducted by Hitch and colleagues (in press) in order to confirm predictions based on simulations generated from a computational model of verbal sequence learning (Burgess & Hitch, 1999, 2006; see section 2.5 for a review of this model and its predictions). These experiments involved manipulating articulatory suppression (Experiments 1 to 3) and phonological similarity (Experiment 3) across different presentation modalities and observing their effects on ISR performance and long-term sequence learning. In Experiment 1, participants were aurally presented with 12-digit sequences, with the repeated sequence presented every second trial. Articulatory suppression was required during presentation and recall. The results revealed an effect of articulatory suppression on ISR performance; more errors were found under conditions of articulatory suppression. However, articulatory suppression failed to have any effect on the Hebb Effect; performance improved over trials for the repeated sequence regardless of whether participants were engaging in articulatory suppression or not. Furthermore, in line with other studies (e.g. Hitch et al., 2005; Schwartz & Bryden, 1971), learning of the repeated sequence was associated with the beginning of sequences.

Experiment 3 investigated the effect of articulatory suppression and phonological similarity when the materials to-be-learned were presented visually. Eight-item letter sequences were used in order to manipulate phonological similarity. The results confirmed and extended those of Experiment 1. Articulatory suppression reduced ISR performance but had no effect on the learning of the repeated sequence. In addition, phonological similarity had its expected effect on ISR performance: a PSE was observed without articulatory suppression, but was abolished under articulatory suppression. Moreover, equivalent rates of learning were observed for the phonologically similar (PS) and phonologically distinct (PD) repeated sequences when articulatory suppression was not required. However, under articulatory suppression, learning of the repeated PS sequence was marginally faster than the learning of the repeated PD sequence. Interestingly, the finding that learning occurred for both PS and PD repeated sequences provided evidence that participants are capable of learning more than one sequence simultaneously. This finding has also been reported by Cumming et al. (2006).

Experiment 2 examined the effect of changing temporal grouping structure over repeated presentations (see section 2.3.5 for details of this experiment). Although engaging in articulatory suppression had the expected effect of reducing ISR performance, it failed to disrupt the learning of the repeated sequence in both RS and RD conditions. The authors interpret this latter finding as evidence that phonological coding is not an important part of long-term sequence learning.

The finding that articulatory suppression and phonological similarity exerted their typical effects on ISR performance confirms that participants were adopting phonological coding of the letter sequences and were therefore utilising the phonological loop. However, the finding that neither of these two variables disrupted the learning of the repeated sequence was taken as evidence that sequence learning is not mediated by the process of subvocal rehearsal, as originally proposed by Cohen and Johansson (1967b), Cunningham et al. (1984), and Kidd and Greenwald (1988), and does not rely on phonological coding. To encapsulate their findings, Hitch et al. (in press) state that "long-term learning resulting from a single trial is not entirely reflected in what can be recalled on that trial (STM)".

Further support for the claim that sequence learning does not depend upon the rehearsal process comes from Fallon et al. (2005). They found reliable sequence learning in conditions where articulatory suppression was required during presentation only, during recall only, and during both presentation and recall. The authors concluded that it is specifically covert rehearsal that plays no role in the Hebb Effect.

Page and colleagues (2006) further investigated the effect of articulatory suppression on the Hebb Effect using visually presented letters (Experiment 1) and pictures (Experiment 2) either without articulatory suppression or under articulatory suppression during presentation and recall. Recall adopted a serial reconstruction procedure, whereby each letter or picture sequence appeared on a computer screen in a "noisy" circle, with a mouse-click as the response. Experiments 1 and 2 reported similar findings: equivalent Hebb Effects were observed with and without articulatory suppression. The factor of phonological similarity was also manipulated in Experiments 1 and 2: in both experiments, a PSE was found for ISR performance when articulatory suppression was not required, but this effect was abolished under articulatory suppression conditions, confirming that articulatory suppression was effective in blocking the use of the phonological loop. In light of the persistence of reliable Hebb Effects despite access to the phonological loop being denied due to articulatory suppression, the authors conclude that the "strong hypothesis that access to the phonological loop is necessary to produce a Hebb repetition effect has clearly been falsified" (p. 720).

In support of Page et al.'s (2006) claim that the Hebb Effect may not be dependent upon the operation of the phonological loop, a number of studies have shown evidence for a Hebb Effect with non-verbal materials (e.g. Corsi, 1972; Couture & Tremblay, in press; Gagnon et al., 2004, 2005; Turcotte et al., 2005). Couture and Tremblay (in press) directly compared a verbal (letter task) and visuospatial (dot task[6]) Hebb repetition task. The magnitude of the Hebb Effects observed in both tasks was equivalent, prompting the authors to postulate that this effect is functionally equivalent across the two domains. This equivalence between the verbal and visuospatial domains is further supported by the observation of numerous empirical similarities between ISR for verbal and non-verbal materials (e.g. Avons, 1998; Jones, Farrand, Stuart & Morris, 1995; Smyth, Hay, Hitch & Horton, 2005; Ward, Avons & Melling, 2005).

Summary

Contrary to earlier studies (e.g. Cohen & Johansson, 1967a, 1967b; Cunningham et al., 1984; Kidd & Greenwald, 1988), it has recently been shown that the Hebb Effect is in fact not mediated by the subvocal rehearsal mechanism (e.g. Fallon et al., 2005; Hitch et al., in press; Page et al., 2006). Although engaging in articulatory suppression reduces ISR performance with both auditory and visual presentation, and eliminates the PSE with visual presentation, articulatory suppression does not prevent the long-term learning of a repeated sequence. These findings suggest that although the phonological loop is utilised when recalling individual sequences, the learning of a sequence does not rely on phonological coding (e.g. Hitch et al., in press; Page et al., 2006). However, the learning of a sequence does appear to be affected by the temporal characteristics of a sequence (e.g. Hitch et al., in press). Studies have also demonstrated that participants are capable of learning more than one repeated sequence simultaneously (e.g. Cumming et al., 2006; Hitch et al., in press). A number of studies also have also provided evidence of Hebb Effects for non-verbal materials (e.g. Corsi, 1972; Couture & Tremblay, in press; Gagnon et al., 2004, 2005; Milner, 1971; Turcotte et al., 2005).

The suggestion that long-term sequence learning does not rely on subvocal rehearsal appears to support the natural process of vocabulary acquisition. Young children are capable of acquiring new words despite not utilising the rehearsal mechanism (e.g. Gathercole, Adams et al., 1994). Furthermore, the finding that

---

[6] Couture and Tremblay (in press) adopted a modified version of the dot task (see Jones et al., 2005) in which a sequence of dots are presented in separate spatial locations on a computer screen for immediate serial recall via a reconstruction method.

participants can learn two sequences simultaneously also represents vocabulary learning; children are capable of learning several new words at once. The finding that functionally similar Hebb Effects occur in the verbal and visuospatial domains (e.g. Couture & Tremblay, in press) may suggest that the mechanism responsible for long-term learning may be a more domain-general one. However, such a finding does not necessarily rule out the importance of the phonological loop in long-term learning. Indeed, it may be that separate learning mechanisms are employed in the verbal and visuospatial domains.

2.3.8 Effect of Age

Given that numerous researchers have recently investigated the Hebb Effect with a view to determining whether it makes use of the phonological loop (e.g. Fallon et al., 2005; Hitch et al., in press; Page et al., 2006), it is perhaps surprising that none of these studies have sought to examine the Hebb Effect in children. This paucity of research is even more remarkable when one considers the potential of this learning paradigm as an experimental tool for investigating the mechanisms underlying the acquisition of new words. That evidence is accumulating supporting the idea that the phonological loop plays a crucial role in the process of language learning (e.g. Baddeley et al., 1998) would lead one to ask the question as to whether the Hebb Effect can be observed in children, especially given that childhood represents an intense period of vocabulary acquisition. Evidence for Hebb Effects in children would support the use of this learning paradigm within the new phonological word-form learning literature. Furthermore, since it is known that children under the age of approximately 4 years have not yet acquired the ability to utilise subvocal rehearsal strategies (e.g. Flavell et al., 1966; Gathercole, Adams et al., 1994; Gathercole & Hitch, 2003), it is of even more interest to determine whether children of this young age show the Hebb Effect.

Fortunately, one such study has been conducted to address just these questions. Hitch and colleagues (Hitch et al., 2006) presented 7- and 11 year old children with sequences of digits to recall with either one (Experiment 2) or two (Experiment 1) filler sequences intervening between each repeated sequence. The length of sequences was adjusted to meet individual children's STM span (Experiment 2). Hebb Effects were observed in both age groups; however, these effects were rather weak in comparison to adult data. Furthermore, despite better ISR performance for repeated compared with filler sequences, no cumulative learning was observed for the repeated sequences. Stronger Hebb Effects were subsequently found when the repeated sequences were presented without any intervening filler sequences (Experiment 3), suggesting that the presence of the filler sequences were disruptive for children of this age. Interestingly, this latter result was extended to 5 year old children, providing further confirmatory evidence that the Hebb Effect does not depend on the process of rehearsal (e.g. Fallon et al., 2005; Hitch et al., in press; Page et al., 2006).

That only weak Hebb Effects were observed in the presence of filler sequences is somewhat reminiscent of Melton's (1963) findings. Hitch and colleagues (2006) relate this finding to the possibility that children may suffer from increased sensitivity to factors such as decay or interference, and that the presence of filler sequences may be more disruptive compared with adults. Alternatively, they suggest that spacing between repeated sequences may generate more forgetting; that is, there may be a balance between learning and forgetting, in that learning of the repeated sequence is subsequently offset by forgetting arising during the process of recalling the filler sequences. They suggest this balance may be different between children and adults. Indeed, Cumming et al. (2006) have already shown that adults are not sensitive to repetition spacing as long as the items within the repeated and filler sequences do not overlap.

Interestingly, the authors propose that using digit sequences constrains sequence learning to arise at the lexical level of representation. In contrast, the process of new phonological word-form learning takes place at the sublexical level of representation, where only phonological information is available. To effectively assess the validity of the Hebb repetition paradigm as a model of new phonological word-form learning, they suggest the paradigm needs to be adapted to involve the learning of sequences of novel word-forms.

2.4 The Hebb Effect and Implicit Serial Learning

Numerous studies reviewed in the preceding sections have commented upon the implicit nature of the learning observed in the Hebb repetition task (e.g. Gagnon et al., 2004; Hitch et al., in press; Page et al., 2006). Several of these studies have shown that the Hebb Effect is not dependent on awareness of sequence repetition (e.g. Hitch et al., in press; McKelvie, 1987; Turcotte et al., 2005). Furthermore, Hitch et al. (in press) report that repetition awareness decreased under conditions of AS, suggesting that sequence learning may be considered a form of implicit serial learning (e.g. Gagnon et al. 2004; Seger, 1994; Stadler, 1993). Studies conducted with patients suffering from neuropsychological damage resulting in impaired memory have shown normal Hebb Effects (e.g. Baddeley & Warrington, 1970; Gagnon et al., 2004). Given these findings, one could argue that the Hebb repetition task reflects implicit serial learning to some extent, although the influence of repetition awareness on the production of the Hebb Effect remains controversial (see Sechler & Watkins, 1991).

Despite the view that the Hebb repetition task may reflect an implicit process, it remains important to highlight the differences between this task and those tasks generally used in the implicit learning literature. The main difference refers to the explicit nature of the recall component involved in the Hebb repetition task: participants in this task are typically asked to recall each sequence immediately following its presentation. Thus, the Hebb Effect may be more accurately characterised as an implicit learning effect within the context of an explicit memory task. In contrast, implicit learning tasks, such as the SRT task, do not involve such an explicit recall component. Learning of a repeated sequence is typically reflected in a faster reduction in reaction times to a specified target compared with filler sequences. Stadler (1993) has directly examined the link between the Hebb Effect and implicit serial learning using a modified version of the SRT task; filler sequences were introduced into the standard SRT task to bring this task more in line with the Hebb repetition task. Hebb Effects were found in the SRT task; moreover, the effect did not depend on participants' ability to recognise sequence repetition. In line with this view, Seger (1994) refers to the Hebb repetition task as reflecting an implicit learning task. The development of more objective measures of assessing repetition awareness may aid in the classification of the Hebb repetition task as an implicit learning task.

2.5 Computational Models of Serial Order and Predictions for the Hebb Effect

Numerous computational models of STM have been developed in order to provide an account of how serial order information is encoded, stored and retrieved (e.g. Brown et al., 2000; Burgess & Hitch, 1992, 1999, 2006; Gupta & MacWhinney, 1997; Gupta, 2006b, cited in Gupta, 2006a; Henson, 1998; Lee & Estes, 1977; Lewandowsky & Murdock, 1989; Page & Norris, 1998). These models attempt to simulate existing behavioural data and generate predictions which can then be tested in order to differentiate between theories of STM. However, as was highlighted in Chapter One, many of these computational models of STM have yet to implement a mechanism for long-term learning. Given that the Hebb repetition paradigm represents a technique for assessing the long-term learning of sequences, the majority of existing computational models are unable to simulate the Hebb Effect. This is disappointing given that this paradigm offers an exciting opportunity to examine the interaction between STM and LTM; indeed, Burgess and Hitch (2005) have recently argued that the "Hebb repetition effect provides a powerful vehicle for developing and testing models of the relationship between STM and LTM" (p. 535). However, there appears to be two computational models which have incorporated mechanisms for long-term learning and serial order (e.g. Botvinick & Plaut, 2006; Burgess & Hitch, 1999, 2006). These models are therefore of particular relevance to the research presented in this thesis. However, only the Burgess and Hitch (1999, 2006) model to date has attempted to simulate the Hebb Effect. Although the recent model proposed by Botvinick and Plaut (2006) claims to simulate the Hebb Effect, this data is yet to be fully reported (Botvinick & Huffstettler, 2006, cited in Botvinick & Plaut, 2006). As a result, the following will provide an overview of the architecture and operationalisation of the Burgess and Hitch (1999; 2006) model, along with its predictions for the Hebb Effect.

Neural Network Model of the Phonological Loop

Burgess and Hitch (1992, 1999, 2006) have developed a simplistic neural network model of the phonological loop which is capable of simulating a wide range of empirical data. For example, the model successfully captures effects of word length, phonological similarity, articulatory suppression and temporal grouping on ISR, along with many of the differing patterns of errors associated with these effects, such as the increase in order errors for phonologically similar material (Conrad, 1965). The model is based on the view that item and order information are separate. This view is represented by the implementation of two main components: a phonological/lexical layer for item information (i.e. phoneme and item information, respectively) and a context/timing signal responsible for the encoding of serial order (see Figure 2.1). These three types of information are represented by different layers of nodes; these nodes transmit activation to nodes in adjacent layers depending on changes in the strength of the connections between them. This strengthening process involves two types of connection weights: one large-amplitude and short-term, and the other small-amplitude and long-term. This first process is responsible for rapid decay and subsequent forgetting, whereas the latter process is slow to decay and permits cumulative learning of repeated inputs, provided not too much time has elapsed between repetitions (Burgess & Hitch, 1999; Hitch et al., in press). Activation of nodes within the context layer represents the timing of item presentation and encodes the order of items via position-item associations.

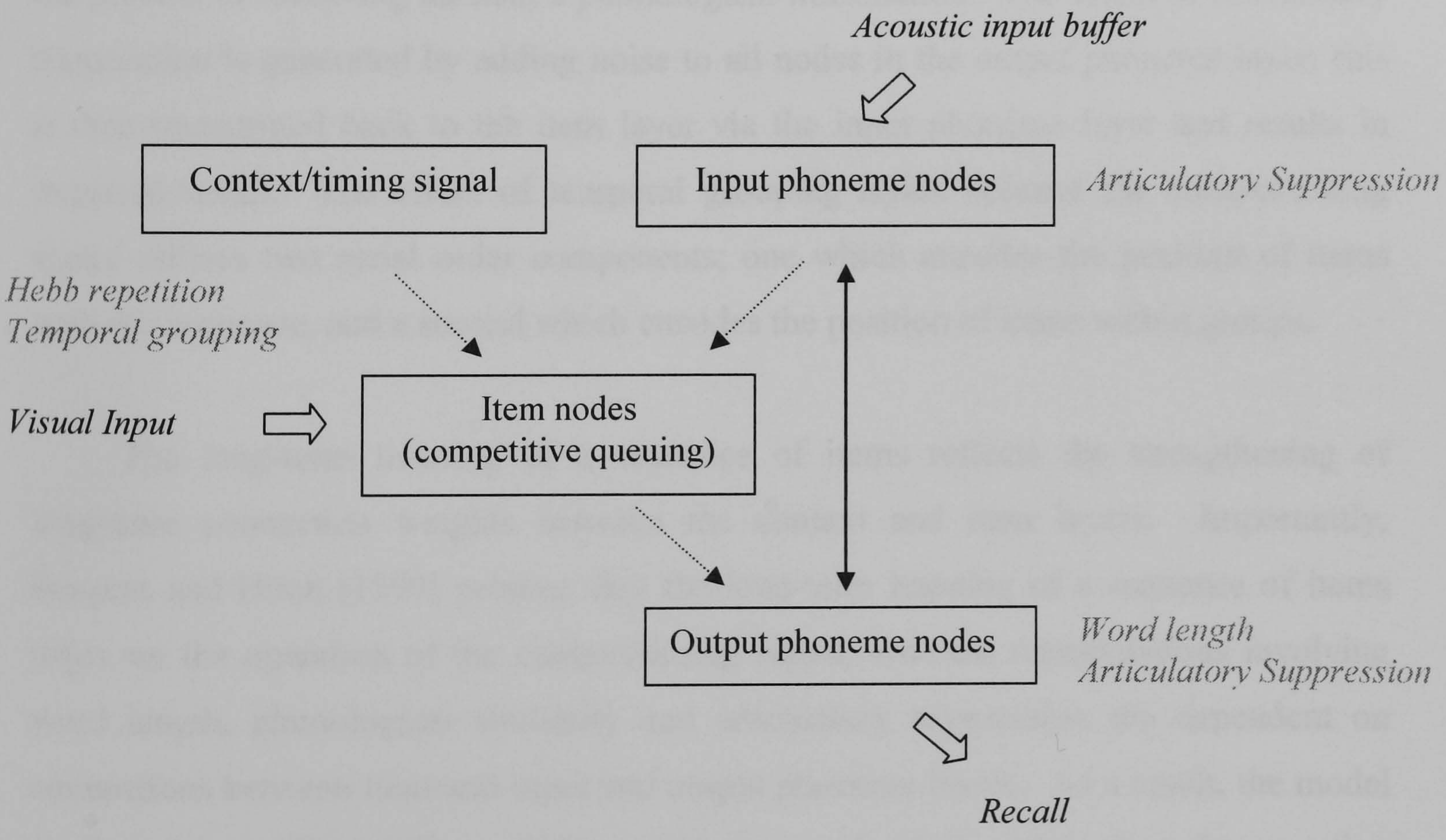


Figure 2.1: *Structure of the Burgess and Hitch model (1999) adapted from Burgess and Hitch (1999, 2005, 2006). Boxes denote layers of nodes and what they represent. Dashed lines represent both short- and long-term modifiable connection weights. Solid lines denote pre-wired one-to-one long-term connections. Blocked arrows represent where auditory and visual inputs access the system. Grey text represents a number of experimental effects captured by the model next to the relevant part of the model.*

During the presentation of a sequence of items, connections between the context and item layers are strengthened along with item-phoneme and phoneme-item connections. The act of recalling a sequence involves a process referred to as 'competitive queuing'. The initial pattern of activation in the context layer spreads to the item layer where the most active item is selected and other items are suppressed. This activation then spreads to the output phoneme layer where the item is recalled; this item is then immediately inhibited to allow selection of the next item. Items selected for output slowly recover from this inhibition to allow for their selection at a later stage. Burgess and Hitch (1999, 2006) highlight the two-stage recall process of this model: each item is processed firstly for serial order information and then for its phonological information (Hitch et al., in press).

The model reproduces the WLE because longer words in a sequence take longer to rehearse and recall, which allows more time for short-term connection weights to decay. The PSE arises due to an increase in 'cross-talk' in phonemic feedback during the process of retrieving an item's phonological information. The effect of articulatory suppression is generated by adding noise to all nodes in the output phoneme layer; this is then transmitted back to the item layer via the input phoneme layer and results in impaired recall. The effect of temporal grouping arises because the context/timing signal utilises two serial order components; one which encodes the position of items with the sequence, and a second which encodes the position of items within groups.

The long-term learning of a sequence of items reflects the strengthening of long-term connection weights between the context and item layers. Importantly, Burgess and Hitch (1999) propose that the long-term learning of a sequence of items relies on the operation of the context/timing signal, whereas manipulations involving word length, phonological similarity and articulatory suppression are dependent on connections between item and input and output phoneme layers. As a result, the model predicts that the Hebb Effect will be sensitive to manipulations that affect the operation of the context/timing signal, such as temporal grouping, but not to manipulations that affect the phoneme layers, such as word length or phonological similarity. The model succeeds in generating data in support of recent behavioural data on the Hebb Effect. It captures the finding that sequence learning is disrupted by changing the temporal characteristics of a sequence at each repetition (Hitch et al., in press), along with the finding that articulatory suppression and phonological similarity impairs ISR performance, but yet fails to disrupt the learning of a repeated sequence (Hitch et al., in press; Page et al., 2006).

The authors have since revised this model (Burgess & Hitch, 2006) to account for behavioural data which suggests that participants are capable of learning more than one repeated sequence simultaneously (e.g. Cumming et al., 2006; Hitch et al., in press). This has been implemented by introducing multiple context-sets. Each sequence presented is matched against previous patterns of associations; if the cumulative match falls below a certain threshold then that set of context nodes is discarded. As successive items are presented, the cohort of context sets gets smaller. If there is a match between context nodes then cumulative learning will occur for that sequence. If there is no match then that sequence recruits a new set of context nodes; this latter process reflects the learning of multiple repeated sequences. Moreover, the idea of multiple sets of context nodes reflects the importance of the start of a sequence (e.g. Hitch et al., 2005; Schwartz & Bryden, 1971); if there is no match between context sets at the beginning of a sequence then cumulative learning will not occur.

The earlier version of the Burgess and Hitch model (1999) predicts Melton's (1963) empirical data that cumulative learning of a sequence reduces as a function of the time elapsed between repetitions; that is, the larger the number of filler sequences presented between repetitions, the smaller the size of the Hebb Effect. However, this version does not predict the findings of Cumming et al. (2006) that repetition spacing has no effect on the size of the Hebb Effect; indeed, these authors observed cumulative learning with repetition spacings of up to 11 filler sequences when different sets of items are presented for the Hebb and filler sequences (i.e. the non-overlapping condition). Cumming et al. (2003) have also tested the prediction that learning a sequence of items involves the strengthening of position-item associations. The earlier version of the Burgess and Hitch model (1999) predicts that the learning of a repeated sequence will transfer to another sequence with a related structure. For example, transfer of learning would be observed if alternate items in the transfer sequence maintain the same serial positions as in the original repeated sequence, resulting in enhanced performance for items at those specific positions. Cumming et al. (2003) found equivalent levels of performance for those items in the transfer sequence that occupied their original position and those that changed serial position. The same pattern of results has also been reported by Hitch et al. (2005). These results suggest that the cumulative learning of a sequence is not a result of strengthening between position-item associations. Cumming et al. (2003) propose instead that the Hebb Effect arises due to the formation of a number of chunks in LTM, with each chunk corresponding to sub-sequences of the entire sequence. However, the recent version of the Burgess and Hitch model (2006) is able to account for Cumming et al.'s (2003) findings by the introduction of multiple context-sets, as described earlier. Cumulative learning of a sequence will occur only if there is a match between context nodes; a new set of context nodes would therefore be recruited for sequences in which items had changed serial position from the original repeated sequence; this would be particularly applicable if these items were located at the beginning of the sequence (e.g. Hitch et al., 2005; Schwartz & Bryden, 1971).

It is clear that Burgess and Hitch's (2006) neural network model of the phonological loop succeeds in capturing the majority of empirical data relating to the Hebb Effect. However, this model is based on the learning of sequences of *familiar* items such as digits, letters or words. If the Hebb repetition paradigm is to be viewed as a potential analogue to the long-term learning of novel phonological word-forms, then the model needs to extend its predictions to the learning of sequences of unfamiliar word-forms. Research conducted utilising novel words as the materials to-be-learned will pose a challenge for models simulating the long-term learning of familiar sequences only.

2.6 Chapter Summary

The studies reviewed in the preceding sections confirm that the Hebb Effect is a robust effect. It has been observed when presenting the materials to-be-learned in both the auditory and visual modalities and with a variety of materials including; digits, letters, words, pseudowords and nameable pictures (e.g. Cohen & Johansson, 1967a; Cumming et al., 2006; Gagnon et al., 2004; Hitch et al., 2005, in press, Melton, 1963; Page et al., 2006). The Hebb Effect also emerges with different types of recall such as; spoken, written and keyboard responses (e.g. Bartz, 1969; Cunningham et al., 1984; Hitch et al., in press; Kidd & Greenwald, 1988; Melton, 1963) and also occurs with serial reconstruction and SRT methods (e.g. Gagnon et al., 2004; Page et al., 2006; Stadler, 1993). The Hebb Effect also extends to the visuospatial domain (e.g. Couture & Tremblay, in press; Gagnon et al., 2004; Turcotte et al., 2005).

A number of factors have been shown to influence the Hebb Effect. Early studies advocate the importance of overt recall and rehearsal in producing the Hebb Effect (e.g. Cohen & Johansson, 1967a, 1967b; Cunningham et al., 1984; Kidd & Greenwald, 1988). However, these studies do not appear to have been individually replicated. In contrast, more recent studies have found that the Hebb Effect is not mediated by rehearsal; the long-term learning of a repeated sequence is not affected by articulatory suppression (e.g. Fallon et al., 2005; Hitch et al., in press; Page et al., 2006). In line with this finding, children as young as 5 years of age appear to show Hebb Effects (e.g. Hitch et al., 2006), although this finding requires further replication. Other studies have found that the Hebb Effect emerges independently of repetition spacing and an interpolated distractor activity, but is instead affected by the degree of item overlap between repeated and filler sequences (e.g. Bartz, 1969; Cumming et al., 2006). The temporal structure of a sequence has been found to play an important role in the learning of a repeated sequence, with the beginning of a sequence proving particularly crucial (e.g. Bower & Winzenz, 1969; Hitch et al., 2005, in press; Schwartz & Bryden, 1971). The role of repetition awareness in the Hebb Effect still remains fairly controversial. Although studies have shown that the Hebb Effect emerges independently of repetition awareness (e.g. Hitch et al., in press; McKelvie, 1987), these results may suffer from subjective and/or retrospective bias (see Sechler & Watkins, 1991). However, studies conducted with amnesic patients may provide support for the idea that the long-term learning of a repeated sequence can proceed at an implicit level, that is, without awareness of sequence repetition (e.g. Baddeley & Warrington, 1970; Gagnon, et al., 2004), although it is important to note that awareness of sequence repetition was only explicitly tested in one of these studies (e.g. Gagnon et al., 2004). Finally, a recent computational model has succeeded in simulating the Hebb Effect under varying circumstances (e.g. Burgess & Hitch, 1999, 2006).

2.7 The Hebb Repetition Paradigm as an Analogue to New Word Learning?

The studies reviewed in the current chapter have provided considerable evidence relating to the factors that contribute to the emergence of the Hebb Effect. Of particular interest are the findings of recent studies suggesting that the Hebb Effect does not depend on the process of subvocal rehearsal or phonological coding (e.g. Hitch et al., in press; Page et al.,2006). However, predictions generated from a computational model of verbal sequence learning (Burgess & Hitch, 1999, 2006) pinpoints the locus of the Hebb Effect to a context/timing signal within the phonological loop which is responsible for encoding serial order. This model accurately predicts that variables known to be dependent upon phonological coding and/or subvocal rehearsal, such as phonological similarity and articulatory suppression, fail to disrupt sequence learning on the basis that these variables selectively utilise the components of the model responsible for processing phonological information, and not serial order information.

The majority of the studies reviewed utilise the Hebb repetition paradigm to assess the long-term learning of sequences of familiar items such as digits, letters and words. Likewise, the Burgess and Hitch model (1999, 2006) is based on the long-term learning of sequences of familiar items. However, very little research has been conducted on the learning of sequences of unfamiliar items, such as novel phonological word-forms. Parallels can be drawn between the learning of a sequence of familiar items and the learning of a novel phonological representation: both require the maintenance of an ordered representation of a string of phonological information. However, in the latter case the strings of phonological information have never before been encountered. This would presumably place heavier demands on the system responsible for maintaining phonological information. It is therefore feasible to suggest that PSTM would represent such a system. The evidence reviewed in Chapter One suggests that PSTM plays a critical role in maintaining novel phonological representations to allow for the formation of more stable and permanent representations in LTM (e.g. Baddeley et al., 1998). Some of this evidence has been derived from studies which use the paired-associate learning task (e.g. Papagno et al., 1991; Papagno & Vallar, 1992). These studies have shown that the learning of word-nonword pairs relies on PSTM, whereas the learning of word-word pairs is assumed to make use of more non-phonological learning codes, such as semantic or lexical codes. In terms of the Burgess and Hitch model (1999, 2006), one could suggest that the learning of a sequence of unfamiliar items would utilise different components of the phonological loop to those used when learning sequences of familiar items. As such, the learning of unfamiliar items may involve the operation of components previously assumed to be independent of the Hebb Effect, such as the components involved in the processing of phonological information.

The current thesis therefore aims to examine whether the Hebb Effect can be extended to the long-term learning of novel phonological word-forms. The predictions investigated are based on two sources of information: paired-associate learning studies and the predictions of the Burgess and Hitch model (1999, 2006). It is important to highlight at this stage that the research reported in this thesis was conducted before the more recent studies of the Hebb Effect were published (e.g. Hitch et al., in press; Page et al., 2006) and is therefore considered novel at the time the research commenced. Indeed, the series of experiments conducted by Hitch et al. (in press) resemble a number of experiments reported in this thesis. Importantly, the current thesis extends the findings of these recent Hebb Effect studies to the learning of sequences of novel phonological word-forms. As a result, this research presents a challenge for the Burgess and Hitch (1999; 2006) model which has, to date, simulated the learning of sequences of familiar items only.

The experimental chapters that follow are intended to extend our current understanding of the mechanisms underlying the learning of novel phonological word-forms. The crux of this research is based on the hypothesis that this essential and complex process is mediated by PSTM. The remainder of this thesis is therefore divided into two parts. The first part attempts to replicate previous findings relating to the effect of phonological similarity on the learning of novel phonological word-forms via a paired-associate task (Chapters Three and Four). The second part aims to extend these findings to the learning of novel word-forms via the Hebb repetition paradigm (Chapters Five and Six).

2.8 Aims of Thesis

The aims of this thesis are as follows:

1. To extend our current understanding of the role of PSTM in the long-term learning of novel phonological word-forms. This will be investigated by:

    (i) Conducting a replication and extension of Papagno and Vallar's (1992) study in which the effect of phonological similarity on the long-term learning of word-word and word-nonword pairs was assessed.

2. To pursue a new avenue of research by examining whether the Hebb repetition paradigm can be viewed as an experimental analogue to the long-term learning of novel phonological word-forms. This will be investigated by:

    (i) Comparing the extent to which phonological similarity affects the long-term learning of a repeated sequence of either words or nonwords.

Chapter Three: Effects of Phonological Similarity on Immediate Serial Recall and Paired-Associate Learning

3.1 Introduction

This chapter addresses the first aim of this thesis by conducting a further test of the hypothesis that PSTM plays a crucial role in the long-term learning of novel phonological word-forms. Chapter One reviewed a large body of evidence in support of the PSTM hypothesis (e.g. Baddeley et al., 1998). This literature is largely governed by developmental studies, the findings from which converge to reveal a strong association between PSTM and vocabulary acquisition (e.g. Gathercole & Baddeley, 1990a; Gathercole et al., 1991a, 1992, 1994; Service, 1992; Service & Kohonen, 1995). These correlation-based findings have been further substantiated by the investigation of experimental word learning abilities in children. Such studies have shown that PSTM contributes to the long-term learning of unfamiliar material, but not to the learning of familiar material (e.g. Gathercole & Baddeley, 1990b; Gathercole et al., 1997; Michas & Henry, 1994; Papagno & Vallar, 1995), suggesting that PSTM constrains the ease with which new phonological representations are acquired (Gathercole et al., 1997).

As highlighted in Chapter One, the majority of studies within the PSTM and new word learning literature have been conducted with children. This is not surprising given that childhood represents an intense period of vocabulary acquisition (Gathercole & Adams, 1993). Although investigating new word learning abilities in children has the potential to more closely represent the natural process of vocabulary acquisition, particularly when learning is assessed by a paired-associate task, important insights can also be gained about the role of PSTM in learning new word-forms by conducting experimental word learning studies with adults. Such studies can capitalise on the wealth of evidence supporting the architecture and operation of the phonological loop (Baddeley & Hitch, 1974; Baddeley, 1986). As described in Chapter One, a number of PSTM variables have been shown to interfere with the operation of the phonological loop in clearly defined ways; these refer primarily to word length, phonological similarity and articulatory suppression (e.g. Baddeley et al., 1975; Conrad, 1964; Conrad & Hull, 1964; Murray, 1967, 1968). The discovery of these PSTM phenomena have provided support for the existence of a time-limited phonological short-term store which maintains phonological information in a speech-based code, and a subvocal rehearsal process which serves to refresh degrading representations in the phonological short-term store in order to prevent their loss from PSTM (Baddeley & Hitch, 1974; Baddeley, 1986). As a result, these experimental phenomena have become the benchmarks of PSTM and are well accounted for in terms of the functioning of the phonological loop.

Given that the effects of word length, phonological similarity and articulatory suppression have been shown to impact on the phonological loop in specified ways, these variables can be utilised to test the claim that PSTM mediates the long-term learning of new phonological word-forms. If the learning of unfamiliar material is indeed mediated by PSTM, then these variables should reveal corresponding effects on the learning of this material; that is, any interference with the operation of the phonological loop should result in difficulties with the long-term learning of new phonological material (Baddeley et al., 1998).

Despite the relative simplicity of the hypothesis that manipulating variables such as word length, phonological similarity and articulatory suppression will impair the learning of unfamiliar material, only two studies to date have specifically tested this hypothesis (e.g. Papagno et al., 1991; Papagno & Vallar, 1992). These studies investigated the effects of articulatory suppression (Papagno et al., 1991), phonological similarity and word length (Papagno & Vallar, 1992) on the paired-associate learning of familiar and unfamiliar material. A similar pattern of results was obtained across each experiment: the manipulated variable disrupted the learning of word-nonword pairs, but not the learning of word-word pairs. These findings provided crucial empirical support for the PSTM hypothesis. The absence of existing lexical or semantic representations to support the learning of nonwords forces participants to rely on temporary representations of their phonological structure in order to establish permanent representations of this material in LTM. In contrast, the learning of familiar material is mediated by the use of non-phonological codes which make use of stored lexical and semantic representations. However, despite their importance within the literature, it appears that neither of the studies conducted by Papagno et al. (1991) and Papagno and Vallar (1992) has been replicated. This is in stark contrast to the many replications of developmental and experimental studies conducted with children (e.g. Gathercole & Baddeley, 1989a; Gathercole et al., 1992, 1997, 1999; Gathercole & Adams, 1993, 1994; Masoura & Gatherole, 1999, 2005; Michas & Henry, 1994; Service, 1992; Service & Kohonen, 1995).

The main aim of the current chapter is to conduct a replication of Papagno and Vallar's (1992) experiment in which phonological similarity was manipulated during a paired-associate learning task. A further aim was to extend Papagno and Vallar's (1992) findings to English participants and materials, given that the original study was conducted with Italian participants and materials. It is important to establish that effects obtained in one language extend to other languages in order to confirm that any effects found are not an artefact of that specific language and/or materials adopted.

The decision to replicate Papagno and Vallar's (1992) experiment in which phonological similarity was manipulated was twofold. Firstly, the PSE has been found to be a highly robust effect (e.g. Logie et al., 1996), especially when material is presented aurally (e.g. Nairne & Kelly, 1999), and has been shown to emerge with adults (e.g. Conrad, 1964; Gathercole et al., 2001; Wickelgren, 1965a) and with young children, provided the material is auditorily presented (e.g. Gathercole et al., 2001; Hulme, 1987). Secondly, the PSE is often taken to reflect the use of phonological coding within the phonological short-term store (e.g. Baddeley, 2003; Gathercole et al., 2001). It is therefore considered that a manipulation of phonological similarity directly assesses the phonological nature of the phonological short-term store, thus providing a direct and controlled test of the PSTM hypothesis

Two experiments are presented in this chapter. Experiment 1 was conducted with the aim of providing confirmatory evidence that the word and nonword material chosen for use in Experiment 2 produced the standard PSE when presented in the format of a typical ISR task. Experiment 2 addresses the claim that PSTM contributes specifically to the learning of unfamiliar material by conducting a replication of Papagno and Vallar's (1992) study.

3.2 Experiment 1: Establishing Phonological Similarity Effects for Immediate Serial Recall of Words and Nonwords

The aim of Experiment 1 was to select a set of words and nonwords that would elicit the PSE in an ISR task prior to using these materials in the paired-associate learning task (Experiment 2). This was particularly important given that previous research has shown mixed results regarding the PSE with nonwords, particularly nonwords rated as being unwordlike or low in associative value (e.g. Fallon et al., 1999; Gathercole et al., 2001; Karlsen, Imenes, Johannesen, Endestad & Lian, 2007; Lian & Karlsen, 2004; Lian, Karlsen & Eriksen, 2004; Lian, Karlsen & Winswold, 2001; Nairne & Kelley, 1999).

Gathercole et al. (2001) found the standard detrimental effect of phonological similarity for CVC nonwords during an auditory ISR task conducted with adults and children. However, no information was provided regarding the wordlikeness of the nonwords used. Conversely, Lian et al. (2001) revealed contrasting effects of phonological similarity with nonwords differing in their long-term associative values during an auditory ISR task. The authors determined the associative value of nonwords based on a novel reaction time assessment, in which participants were required to provide an associated word or statement to the presentation of individual nonwords. Nonwords were identified as having either high or low associate value based on a median split of participants' reaction times.

Lian et al. (2001) obtained the typical PSE with words and nonwords rated as high in associative value. However, a reverse PSE was found for nonwords rated as low in associative value; that is, phonologically similar nonwords were recalled significantly better than phonologically distinct nonwords. The authors interpreted their findings in terms of the degree of activation of LTM mechanisms, claiming that words and nonwords of high associative value activate existing lexical representations in LTM, whereas nonwords of low associative value fail to activate existing LTM representations. The authors further suggest that Gathercole et al.'s (2001) finding of a PSE with nonwords is most probably due to their degree of wordlikeness, suggesting that Gathercole et al. (2001) used nonwords of high associative value.

Experiment 1 therefore presented an ISR task in which the phonological similarity of aurally presented words and nonwords was manipulated. Nonwords selected for use in Experiment 1 were rated as low in wordlikeness. On this basis, it was predicted that a PSE would be obtained with words and nonwords. Furthermore, a lexicality effect was also predicted such that words would be better recalled than nonwords (e.g. Hulme et al., 1991, 1995).

3.2.1 Method

3.2.1.1 Design

Participants undertook an ISR task in which lists of phonologically similar and phonologically distinct words and nonwords were auditorily presented; participants were then required to immediately recall the items in their correct serial order. The experiment used a 2 x 2 x 2 mixed-design incorporating two within-subject factors: lexicality (words, nonwords) and phonological similarity (similar items, distinct items); and one between-subjects factor: materials (materials A, materials B). The two within-subjects factors combined to produce four experimental conditions: similar words; distinct words; similar nonwords; and distinct nonwords. Participants also completed a pre-test prior to the ISR task; this involved repeating back the experimental items to be presented in the ISR task to check participants could pronounce these correctly.

For the ISR task, the order of presentation of the lexicality and phonological similarity factors was counterbalanced across participants so that half received the word lists followed by the nonwords lists; this order of testing was reversed for the remaining participants. Furthermore, half of the participants received the similar items followed by the distinct items and the remaining half received the reverse. For the pre-test, half the participants received the word lists followed by the nonword lists, and half received the reverse. To increase internal validity, thereby promoting generalisation of the results, two closed sets of materials were created for each of the four conditions. Participants were randomly divided so that half received materials A, and half received materials B. For the ISR task, for each of the four conditions, 15 experimental lists were created for each of materials A and B (i.e. 30 lists for each condition). For each set of materials, the order in which the words or nonwords were presented within each of the 15 lists was pre-randomised and remained the same for each participant; the order of list presentation also remained the same for each participant. For the pre-test, the similar and distinct items were randomly intermixed (for words and nonwords separately) and the order in which these items were presented remained the same for each participant.

3.2.1.2 Participants

Twenty-four undergraduate students attending the University of York volunteered to take part for either course credit or payment. There were 21 females and 3 males aged between 18 years and 24 years (mean age of 20.1 years). Participants all spoke English as their native language and reported no known hearing or language impairments.

3.2.1.3 Apparatus

The conditions were delivered on an iMac OS 9 computer which operated the required software in SoundEdit via Hypercard version 2.4. Participants heard the materials via a pair of Sennheiser headphones. A Sony portable minidisc recorder MZ-N710 was used in order to record the materials, and to record and play back participants' responses.

3.2.1.4 Materials

The materials were all two-syllable words and nonwords. Each experimental list contained either five words or four nonwords; the smaller size of the nonword lists was an attempt to equate task difficulty. Each experimental list was selected from 8-item sets.

Word Sets

Four sets of eight words were constructed: two phonologically similar sets (materials A and B) and two phonologically distinct sets (materials A and B); see Table

3.1. Each word began with a single consonant and incorporated two vowels, one in each syllable (e.g. *bullet, content, recall*). Each of the four sets were matched on written frequency per million (Kucera & Francis, 1967; mean range across sets: 30 to 36) and concreteness (mean range across sets: 4.51 to 4.82) using the MRC Psycholinguistic Database (Coltheart, 1981). A number of two-tailed independent t-tests revealed no statistically significant differences between any of the four word sets for either concreteness or written frequency (all *p*s>.05). See also Appendix 1(a) for these word sets and their properties.

Table 3.1: *Phonologically Similar and Distinct Word Sets*

Phonologically Similar Word Sets		Phonologically Distinct Word Sets	
Materials A	Materials B	Materials A	Materials B
suspect	pocket	narrow	pupil
supper	pollen	lumber	rubber
sunset	powder	delight	safety
hunter	copper	parish	custom
hunger	content	recall	mortar
budget	concert	mustard	wedding
bullet	wonder	column	highway
burden	worker	bishop	fellow

Phonologically Similar Word Sets: These were constructed so that the words had similar sounding phonological representations (but did not all rhyme). Within a set, each of the eight words had a similar first vowel and a similar second vowel (e.g. *suspect, hunter*). Three different initial consonants and four (materials A) or five (materials B) different final consonants were used within each set.

Phonologically Distinct Word Sets: These were constructed so that the words had distinct sounding phonological representations. Within each set, each word had a different ordering of vowels for the first and second vowel positions (e.g. *pupil, custom*). Different initial and final consonants were used for each of the eight words within a set.

Nonword Sets

A cohort of 136 nonwords was devised. Each of the nonwords conformed to the same rules that were applied when selecting the words. Thus, each nonword began with a single consonant and had one vowel in each syllable (e.g. *bergops, jorlam, suttic*). Participant ratings of the degree of wordlikeness of each nonword were obtained to allow for the selection of those nonwords with the lowest wordlikeness ratings.

For the wordlikeness ratings task, each nonword was presented via PC speakers to either individual or pairs of participants in a quiet testing room. Following Martin and Gathercole's (1997, cited in Gathercole et al., 1997) procedure, participants were advised to make a judgement based on the degree to which the spoken form of each nonword would "pass for a real word in the English language" using a five point rating scale. A score of 1 denoted a nonword that would be '*very unlikely to pass for a real word in English*', and a score of 5 indicated a nonword that would be '*very likely to pass for a real word in English*'. Participants were given a maximum of 7 s to make their responses by ticking the appropriate box (from 1 to 5) on a response sheet.

Twenty-three participants from the University of York volunteered to take part in this task for either course credit or payment. Participants were a mixture of undergraduates and postgraduates. There were 5 males and 18 females aged between 18 years and 30 years (mean age 20.1 years). All participants spoke English as their native language and had no known hearing impairments.

Four sets of eight nonwords were chosen from those nonwords with the lowest wordlikeness ratings: two phonologically similar sets (materials A and B) and two phonologically distinct sets (materials A and B); see Table 3.2. Each of the sets was constructed following the same criteria as were adopted for the construction of the phonologically similar and distinct word sets. The nonword sets were matched on mean wordlikeness ratings (mean range across sets: 1.9 to 2.3). A number of two-tailed independent t-tests revealed no statistically significant differences in wordlikeness ratings between any of the four nonword sets (all *p*s>.05). See also Appendix 1(b) for the nonword sets and their properties.

Table 3.2: *Phonologically Similar and Distinct Nonword Sets*

Phonologically Similar Nonword Sets		Phonologically Distinct Nonword Sets	
Materials A	Materials B	Materials A	Materials B
merglip	paddip	bamich	cepfil
mefflib	paglip	lebbist	butkels
memblin	darglit	tafflost	jorlam
feppip	dasklint	musglent	sibbart
feggin	damklin	cuddow	rodgunt
febslib	lappish	pevtong	welptar
bebbict	larmip	dapeth	fiddop
berpict	labblin	suttic	ludgash

Phonologically Similar Nonword Sets: Each of the eight nonwords had a similar first vowel and a similar second vowel (e.g. *berpict*, *memblin*). Three different initial consonants and four (materials A) or five (materials B) different final consonants were used.

Phonologically Distinct Nonword Sets: Each nonword had a different ordering of vowels for the first and second vowel positions (e.g. *rodgunt*, *fiddop*). Different initial and final consonants were used for each of the eight nonwords within a set.

Construction of Experimental Lists

Phonologically Similar & Distinct Words: Each of the 15 similar and 15 distinct 5-item lists was randomly generated from the corresponding 8-item sets, for materials A and B separately. The only criteria adopted were that words which may be perceived to have strong semantic links (e.g. *hunger-supper*; *bishop-parish*) only appeared in the same list on a few occasions (four or five) and when they did they had a minimum of two words separating them. Of the eight similar and distinct words, five appeared in nine lists and three appeared in ten lists (for both materials A and B).

Phonologically Similar & Distinct Nonwords: Each of the 15 similar and 15 distinct 4-item lists was randomly generated from the corresponding 8-item nonword sets, for materials A and B separately. The selection of these lists was randomised with no constraints. Of the eight similar and distinct nonwords, four appeared in seven lists and four appeared in eight lists (for both materials A and B).

Editing of Speech Signals

Each item was recorded in a soundproof room by a female native English speaker onto a minidisc player. These speech sounds were then transferred to a computerised SoundEdit software package in 16-bit amplitude format. The sampling rate was set to 44,100 kHz and the items were edited so as to normalise their amplitude profiles. Each item was recorded separately to avoid the possibility of co-articulation. Periods of silence were inserted into each sound file so that when the items were played one after another (in the ISR task) they appeared to occur at equal intervals.

3.2.1.5 Procedure

The ISR task was conducted with individual participants in a quiet testing room. Participants were informed the experiment would assess their short-term memory for lists of words or nonwords and would take approximately 35-40 mins. Participants were advised that they would hear several lists containing either five words or four nonwords which they would be required to immediately recall aloud in exactly the same order in which they were presented. Participants were instructed to say 'blank' if they could not remember the item that appeared in a particular position in the list, and to indicate when they were unable to recall any further items from the list. Participants were also advised that before the task commenced, they would hear each item individually and would be required to repeat it aloud. Any items incorrectly articulated were corrected by the experimenter and participants were to repeat each of these items until they had been correctly pronounced. Participants were made aware that their responses would be recorded on to a minidisc player in order to facilitate scoring and so they were to try and speak their responses as clearly as possible.

Two practice lists, one containing five words and the other four nonwords (these items were not in the experimental lists), were given prior to testing the experimental lists. This was in order to check the volume level and to allow participants to become accustomed to the procedure and the speaker's voice.

For the pre-test, each of the items were individually presented and the experimenter made a written record of participants' responses. For the ISR task, each experimental list began with the presentation of a 2 s pause, in which a short beep was presented, followed by the five words or four nonwords. Items were presented at a rate of one per 1.5 s. After the presentation of the final item within a list, a 2 s pause, again including a small beep, was presented indicating the beginning of the spoken recall period. When participants indicated that they were unable to recall any further items from the list, the next list was presented. A 2 min interval occurred between each of the experimental conditions (i.e. after presentation of each set of 15 lists).

Participants' responses were scored both immediately using a strict serial recall criterion and later using the minidisc recordings to verify scoring and to transcribe the exact pronunciations of the items, particularly the nonwords. Maximum mean scores were 5 and 4 for each of the word and nonwords lists, respectively (raw scores of 75 and 60, respectively).

3.2.2 Results

Responses were scored following a strict serial recall criterion; items were scored as correct only if they were phonologically correct and recalled in the correct serial position. Mispronounced items were scored as incorrect. However, if a mispronunciation was consistently made this was scored as correct, provided it was recalled in the correct serial position, on the basis that the item had been misheard but not successfully corrected by the experimenter in the pre-test. This situation was rare occurring for only one participant's responses[7]. Any instances in which the correct pronunciation of an item could not be clearly determined were scored as correct, provided it was recalled in the correct serial position; this situation occurred for only two participants' responses[8].

---

[7] The nonword '*feppip*' was repeatedly mispronounced as '*febbib*'.

[8] The nonword '*cepvil*' was a feasible pronunciation of the nonword '*cepfil*'; and the nonword '*memblim*' was a feasible mispronunciation of the nonword '*memblin*'.

The mean scores obtained from each participant were calculated for each of the four conditions. The data were normally distributed and no outliers were detected[9]. Inter-rater reliability analysis was conducted by an independent rater[10] on a random sample of 15% (n=4) of the participants' responses from the ISR task. This revealed perfect correspondence in terms of serial recall scores for each of the four conditions. The means and standard deviations were calculated for materials A and materials B separately. Percentage scores were also calculated to allow for comparisons between the word and nonword conditions.

The pattern of results demonstrated that words were better recalled than nonwords, irrespective of their phonological similarity; and that distinct items were recalled better than similar items, regardless of their lexical status. This suggests the presence of both a lexicality effect and PSEs. Moreover, an identical pattern of results, coupled with similar levels of performance, was observed for materials A and B, indicating that these effects can be generalised to more than one set of materials.

Participants' percentage scores were entered into a three-way mixed-design analysis of variance (ANOVA) incorporating two within-subject factors: lexicality (words, nonwords) and phonological similarity (similar items, distinct items); and one between-subjects factor: (materials A, materials B). The ANOVA revealed statistically significant main effects of lexicality ($F(1,22) = 741.41$, $MSe = 68.032$, $p<.0001$, $r=.99$) and phonological similarity ($F(1,22) = 71.03$, $MSe = 66.918$, $p<.0001$, $r=.87$), confirming better recall for words compared with nonwords, and for distinct items compared with similar items. All interactions failed to reach significance (all *p*s>.05). Figure 3.1 shows the mean scores for each of the four conditions when combining the data from materials A and B. The size of the PSE is comparable across words and nonwords (difference of 14% for words, and 15% for nonwords).

---

[9] Outliers were defined a priori as scores with a standardised z-score of ≥2.5 away from the sample mean, and were calculated for each condition separately for each of materials A and B separately.

[10] This was a female native English speaker who had not taken part in either Experiments 1 or 2. Of the four sets of participants' responses scored, two had been presented with materials A, and two with materials B.

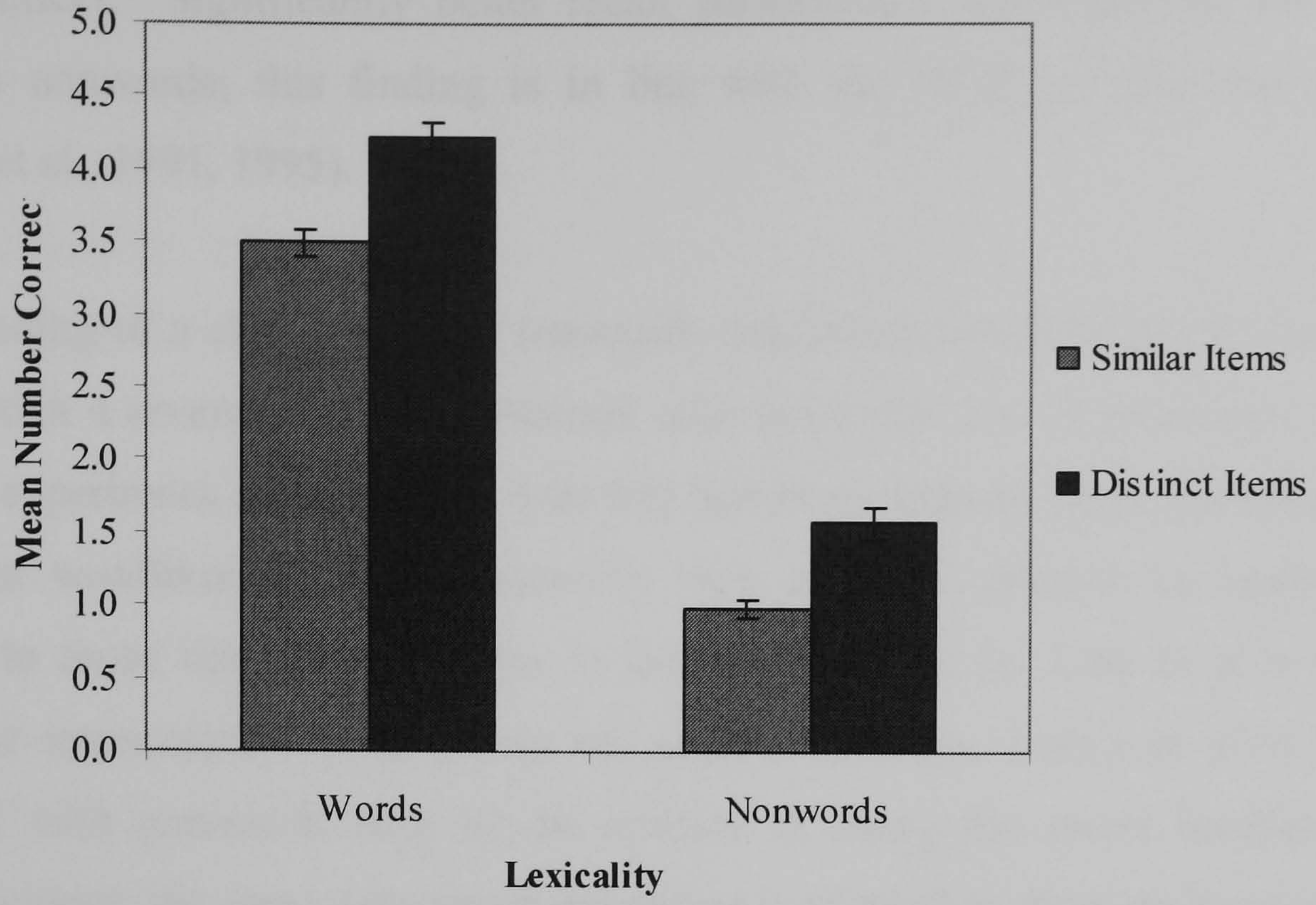


Figure 3.1: *ISR performance as a function of lexicality and phonological similarity (error bars represent standard error of the mean). Maximum mean scores: words = 5; nonwords = 4)*

3.2.3 Discussion

The aim of the current experiment was to confirm that the criteria adopted when selecting the words and nonwords to be used in Experiment 2 would generate the standard negative effect of phonological similarity in an auditory ISR task. Moreover, it was particularly important to clarify the nature of any PSE observed with nonwords in light of recent evidence suggesting a reversal of the PSE with nonwords of low associative value (e.g. Lian et al., 2001). It was predicted that the standard PSE would be observed with both words and nonwords. It was also further hypothesised that evidence for a lexicality effect would be found.

The results yielded clear PSEs for both words and nonwords thereby supporting the first prediction; phonologically distinct items were recalled significantly better than phonologically similar items, regardless of their lexical status. These findings provide confirmatory evidence that the criteria applied to the construction of the phonologically similar and distinct words and nonwords were sufficient to produce the standard detrimental effect of phonological similarity. Moreover, the lack of a lexicality x phonological similarity interaction demonstrates that the size of the PSE was comparable for words and nonwords. The present results also provide clear evidence of a lexicality effect. Significantly better recall performance was observed for words compared to nonwords; this finding is in line with the results of previous studies (e.g. Hulme et al., 1991, 1995).

The finding of a clear PSE with nonwords contradicts the findings of Lian et al. (2001) in which a reverse PSE was obtained with nonwords low in associative value. The present experiment selected nonwords that had been rated by adult participants as being low in wordlikeness; these nonwords may therefore provide an appropriate comparison to those deemed to be low in associative value by Lian et al.'s (2001) reaction time assessment. Considering the present findings, Lian's et al.'s (2001) reverse PSE with nonwords may be an artefact of using the novel reaction time assessment; indeed, the same assessment procedure is adopted in other studies reporting the reverse PSE with nonwords (e.g. Karlsen et al., 2007; Karlsen & Lian, 2005; Lian et al., 2004).

Alternatively, the PSE obtained with nonwords in the present experiment may in part be due to the use of phonologically similar nonrhyming word and nonword lists. This interpretation is in line with Fallon et al.'s (1999) claim that the PSE is restricted to similar nonrhyming lists. Nairne and Kelly (1999) also suggest that using a closed set of items reduces the role of item memory in ISR, as there are less unique items to recall, and so increases the magnitude of the PSE. In the current experiment, there was item repetition across trials for both words and nonwords which may have contributed to the large PSEs observed for both types of material. However, Lian et al. (2001) also used a closed set of items, suggesting that their reverse PSE for nonwords of low associative value was not attributable to reducing the role of item memory. Although these explanations are of interest, the current experiment was not conducted with the intention of investigating the effects of varying the type of phonological similarity used or the way in which ISR performance is scored (i.e. ISR performance for item vs. order).

The fact that clear PSEs were found for ISR of both words and nonwords in the present experiment confirms that phonological similarity was adequately manipulated when constructing these materials; these materials were therefore considered appropriate for use in Experiment 2. This allows a cleaner interpretation of Experiment 2's results; that is, any lack of an effect of phonological similarity on the learning of familiar or unfamiliar material cannot be attributed to an insensitivity of this material to this manipulated variable.

3.3 Experiment 2: Effects of Phonological Similarity on Paired-Associate Learning – A Replication and Extension of Papagno and Vallar (1992)

The aim of Experiment 2 was to replicate and extend Papagno and Vallar's (1992) study in which the effect of phonological similarity on the long-term learning of word-word and word-nonword pairs was investigated. Papagno and Vallar (1992) actually conducted two experiments examining the effects of phonological similarity on paired-associate learning. The findings from their first experiment revealed an effect of phonological similarity on the learning of both nonword- and word pairs, suggesting that PSTM is utilised during the learning of both unfamiliar and familiar material. The finding that phonological similarity disrupted the learning of word pairs contradicted the results reported by Papagno et al. (1991) in which articulatory suppression failed to interfere with the learning of word pairs. Indeed, the results of Papagno and Vallar's (1992) first experiment is in contrast with a number of more recent studies which have shown that PSTM does not contribute to the learning of word-word pairs in children (e.g. Gathercole & Baddeley, 1990b; Gathercole et al., 1997; Masoura & Gathercole, 2005). Papagno and Vallar (1992) argued that their result reflected participants' use of short-term phonological storage rather than non-phonological learning codes, such as semantic codes, during the learning of word pairs.

The authors therefore conducted a second experiment in which a digit span task was interpolated between the presentation and recall phases of the paired-associate task. Papagno and Vallar (1992) predicted that the inclusion of this task would prevent participants from utilising PSTM when learning word pairs, instead inducing them to rely on non-phonological learning codes, as initially predicted. It was hypothesised that this revised method would eliminate the effect of phonological similarity on word pair learning. In contrast, the inclusion of this task was not expected to rule out the use of PSTM when learning nonword pairs. Given that no pre-existing lexical-semantic representations were available for these nonwords, participants would be forced to rely upon phonological coding and storage. The results conformed to Papagno and Vallar's (1992) predictions. Phonological similarity disrupted the learning of nonword pairs, but not the learning of word pairs. This subsequent finding was taken as evidence that PSTM contributes specifically to the learning of unfamiliar material, and that the learning of word pairs can successfully proceed by reliance on non-phonological learning codes.

Experiment 2 therefore replicated Papagno and Vallar's (1992) second experiment but with two notable exceptions. Firstly, the materials used were derived from the English language. The second modification involved controlling for the type of nonwords used. Papagno and Vallar (1992) make no reference to the type of nonwords they used, simply stating that "nonwords were constructed by changing one letter of a real word" (p. 50). In contrast, the present experiment carefully controlled for this factor by only selecting nonwords which had been subjectively rated as being unwordlike (as described in Experiment 1). It was anticipated that using unwordlike nonwords would reduce the opportunity for facilitation from the existing structure of the English language (e.g. Gathercole et al., 1991a; Gathercole, 1995a; Grant et al., 1997; Vitevitch et al., 1997; von Bon et al., 1997; see Chapter One), therefore placing heavier demands on PSTM by forcing participants into relying primarily on the phonological structure of the nonwords. In turn, it was hypothesised that this would elicit a strong PSE during the learning of nonword pairs.

On the basis of Papagno and Vallar's (1992) results, it was predicted that phonological similarity would selectively impair the learning of nonword pairs; that is, significantly faster learning was expected for distinct compared to similar nonword pairs. In contrast, it was predicted that phonological similarity would fail to have any effect on the learning of word pairs, resulting in equivalent learning rates for similar and distinct word pairs.

3.3.1 Method

3.3.1.1 Design

Participants undertook a paired-associate learning task in which they were required to learn eight auditorily presented cue-target pairs, recalling the target item upon presentation of the associated cue item. A digit span task, in which participants recalled a sequence of eight digits in their correct serial order, was interpolated between the presentation and recall phases of the paired-associate task. The pairs were presented over the course of five trials in order to investigate rates of learning. The cue items were all words, and the target items were phonologically similar and phonologically distinct words or nonwords. The experiment used a 2 x 2 x 5 x 2 mixed-design incorporating three within-subjects factors: lexicality (words, nonwords), phonological similarity (similar items, distinct items) and trials (1 to 5); and one between-subjects factor: materials (materials A, materials B). The lexicality and phonological similarity factors combined to produce four experimental conditions: similar word-word pairs; distinct word-word pairs; similar word-nonword pairs; and distinct word-nonword pairs.

The order of presentation of the lexicality and phonological similarity factors was counterbalanced across participants so that half received the word-word pairs followed by the word-nonword pairs, and half received the reverse. Furthermore, half of the participants received the similar cue-target pairs followed by the distinct cue-target pairs and the other half received the reverse. As in Experiment 1, participants were randomly divided so that half received materials A, and half received materials B. For each set of materials in each of the four experimental conditions, the cue-target pairs were presented in a different random order on each of the five trials at both learning and recall phases, with the only constraint being that the last cue-target pair presented in the learning phase was not the first pair to be tested at the recall phase. This order remained the same for each participant.

3.3.1.2 Participants

Twenty-four undergraduate students attending the University of York volunteered to take part for either course credit or payment. There were 22 females and 2 males aged between 18 years and 36 years (mean age of 20.6 years). Participants all spoke English as their native language and reported no known hearing or language impairments.

3.3.1.3 Apparatus

The same apparatus were used as in Experiment 1.

3.3.1.4 Materials

All cue items were two-syllable words. The target items were the materials used in Experiment 1. Each cue-target set contained eight pairs of cue-target items.

Cue Word Sets

Eight sets of eight cue words were constructed: two sets (materials A and B) for each of the four experimental conditions. The cue words were all nouns and were phonologically distinct sounding from each other within each set (e.g. *arrow, duty, garden, thunder, football, navy, talent, weapon*). Each of the eights sets were matched for written frequency per million (Kucera & Francis, 1967; mean range across sets: 27 to 41) and concreteness ratings (mean range across sets: 4.18 to 5.54) using the MRC Psycholinguistic Database (Coltheart, 1981). Two separate one-way between-subjects ANOVAs revealed no statistically significant differences between the eight cue word sets for written frequency ($F(7,56) = 0.79$, $MSe = 322.047$, $p>.05$) and concreteness ($F(7,56) = 1.48$, $MSe = 13772.935$, $p>.05$). See Appendix 2(a) for the cue word sets.

Cue-Target Pairs

Cue word sets were randomly paired with target item sets. The word-word pairs were randomly paired although any obvious associations between the two words were avoided (e.g. *cowboy-bullet*) in order to reduce the ease with which participants could generate strategies to facilitate their memory for these pairs. Each item within a pair had different initial and final consonants (e.g. *cottage-recall*) and did not have a similar phonological structure (i.e. the items did not sound alike or rhyme). See Appendix 2(b) for the cue-target pair sets.

Editing of Speech Signals

Each item was recorded onto a minidisc player in a soundproof room by the same speaker as used in Experiment 1. The cue words were edited following the same procedure as described in Experiment 1.

3.3.1.5 Procedure

The paired-associate task was conducted with individual participants in a quiet testing room. Participants were informed that the experiment involved learning sets of eight word-word and eight word-nonword pairs and would take approximately 60 mins. Each trial would be split into three phases: a listening phase; a digit recall phase; and a pair recall phase[11]. In the listening phase, participants would hear each of the eight pairs of items. The digit recall phase involved the presentation of a single sequence of eight digits (using the digits 1 to 9). Participants were allowed 10 s to recall aloud as many digits as possible in the order in which they were presented. Finally, the pair recall phase involved participants recalling the second member of each pair in response to the presentation of the first member of that pair, allowing 7 s for each pair. Participants were informed these three phases made up a single trial, and five of these trials would make up one set. Furthermore, the pairs would be presented and tested in a different random order on each of the five trials within each set. Beeps were used throughout to signal the beginning and end of each phase. Participants were made aware that their responses would be recorded on to a minidisc player and so they were to try and speak their responses as clearly as possible. Participants were told whether to expect word-word or word-nonword pairs prior to the presentation of each set.

There were two practice trials, one containing eight word-word pairs and the other eight word-nonword pairs. None of these pairs was in the test sets.

Each pair in the listening phase was presented at a rate of 2.75 s with a 2 s interval between each pair, followed by a short beep. Following presentation of the last pair, a longer beep signalled the start of the digit recall phase. The digit sequence was presented at a rate of 1 s per digit; the eighth digit was followed by a short beep indicating the beginning of the digit recall period. The start of the pair recall phase was signalled with a longer beep. The first cue word was then presented, followed by the recall period which ended with a short beep. After presentation of the eighth cue word

---

[11] Experiment 2 did not include a pre-test in which participants were required to listen and repeat back each item individually as in Experiment 1. The pre-test was included in Experiment 1 as the nonwords in the ISR task were presented every 1.5 s, therefore making the task of correctly hearing and repeating these more difficult for participants. This was judged not to be the case in Experiment 2 as a longer time interval of 2 s was used between presentations of each cue-target pair.

and its associated recall period, a longer beep signalled the end of the trial. There was a 30 s pause between each of the four experimental conditions.

Participants' responses were scored both immediately and later using the minidisc recordings for verification and to transcribe the exact pronunciation of the target items, particularly the nonwords, as in Experiment 1. The maximum score was 8 for each of the five trials within a set, with a total score of 40 for each of the four conditions.

3.3.2 Results

3.3.2.1 Correct Recall Analysis

Responses were scored as correct if they were phonologically correct and associated with the correct cue word. Mispronounced items were scored as incorrect. However, if a mispronunciation was consistently made in response to the same correct cue word this was scored as correct on the assumption that the item had been misheard[12]; this situation occurred on five occasions[13]. There were no instances in which the correct pronunciation of an item could not be clearly determined.

The data were not normally distributed. The word conditions were negatively skewed and the nonword conditions were positively skewed at Trial 1[14]. This skewness is presumed to reflect ceiling and floor effects, respectively. Thus, a cautionary note needs to accompany the outcome of the following statistical analyses, particularly those conducted on the word data, and their subsequent interpretation. One participant's data from materials A proved to be an outlier and so this participant's entire data set was excluded from all further statistical analyses[15]. Inter-rater reliability analysis was conducted by an independent rater[16] on a random sample of 15% (n=4) of participants'

---

[12] Consistently mispronounced items were only scored as correct if they occurred on three or more consecutive trials, which had to include the final trial (i.e. on at least trials 3, 4 and 5); in addition, the trials preceding the first instance of the mispronounced item had to have been responded to with a 'blank' response or the misprounced response.

[13] The nonword *'ludgash'* was mispronounced to on two occasions as *'bludgash'* and on one occasion as *'ledgash'*; the nonword *'lebbist'* was mispronounced on one occasion as *'libbist'*; and the nonword *'merglip'* was mispronounced *'mergrin'* on one occasion.

[14] Skewness z-score values were greater than 2.58 (significant at $p<0.01$) based on Field (2005).

[15] Outliers were defined and calculated as in Experiment 1.

[16] The independent rater was the same as in Experiment 1. Of the four sets of participant responses scored, two had been presented with materials A and two with materials B.

nonword responses[17]. This revealed inter-rater reliability estimates of r=1.00 for similar nonword pairs and r=.98 for distinct nonword pairs. The means and standard deviations were therefore calculated based on the 11 participants receiving materials A and the 12 participants receiving materials B.

A reasonably parallel pattern of learning was found across materials A and B, indicating the generalisability of the results. Learning was observed over trials within each of the four conditions. However, the rate of learning was restricted in both word conditions due to near-ceiling performance from Trial 2 onwards. In contrast, performance at Trial 1 in both nonword conditions was near-floor, thus allowing for cumulative learning by Trial 5. The results also demonstrated better learning for word pairs compared to nonword pairs, irrespective of their phonological similarity. However, differential effects of phonological similarity were observed for the word and nonword data. The similar and distinct word pairs appear to be learned to comparable degrees across each of the five trials. This is suggestive of a lack of an effect of phonological similarity on word pair learning. For the nonword conditions, the distinct nonword pairs were learned better at each of the five trials. Moreover, distinct nonword pairs were acquired at a faster rate over trials compared to similar nonword pairs, suggesting that phonological similarity differentially affected nonword pair learning.

Due to clear ceiling effects in both word conditions, and near-floor performance at Trial 1 in both nonword conditions, the word and nonword data were analysed separately. Participants' scores were entered into two separate three-way mixed-design ANOVAs each incorporating two within-subjects factors: phonological similarity: (similar items, distinct items) and trials: (1 to 5); and one between-subjects factor: materials (materials A, materials B). For all analyses reported in this thesis, wherever the assumption of sphericity was violated, degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity, unless indicated otherwise.

The ANOVA conducted on the word data violated the assumption of homogeneity of variances as indicated by significant Levene's tests; variances were significantly

---

[17] Participants' responses in both word pair conditions were not recorded onto minidisc as this was felt unnecessary at the time of testing as the experimenter was confident of accurately reporting these. Inter-rater reliability estimates were therefore only obtained for the nonword pair conditions. However, in all future experiments, minidisc recordings were obtained for all responses made.

different for similar word pairs at Trials 3 and 4 (both *ps*<.05) and for distinct word pairs at Trials 2 and 3 (both *ps*<.05). These significant differences in variances presumably reflect ceiling performance at these trials; thus any slight deviation from this high level of performance is likely to cause large differences in variance. Consequently, it was considered inappropriate to report the statistical results of this ANOVA. However, given that performance at Trial 1 was not at ceiling, a 2 x 2 mixed-design ANOVA was conducted on the word data taking into account the learning of word-word pairs at this single trial only. The ANOVA incorporated one within-subjects factor: phonological similarity (similar words, distinct words); and one between-subjects factor: materials (materials A, materials B).

The ANOVA revealed no statistically significant main effects of phonological similarity (*F*(1,21) = 3.49, *MSe* = 1.033, *p*=.08) and materials (*F*(1,21) = 1.20, *MSe* = 5.483, *p*=.29). The phonological similarity x materials interaction failed to attain significance (F(1,21) = 0.13, MSe = 1.033, p=.73). This confirms an identical pattern of results across materials A and B at Trial 1. Figure 3.2(a) presents the results over all five trials when collapsing over the two sets of materials.

The ANOVA conducted on the nonword data reported statistically significant main effects of phonological similarity (*F*(1,21) = 15.21, *MSe* = 4.769, *p*=.001, *r*=.65) and trials (*F*(1.974,41.446) = 103.51, *MSe* = 2.941, *p*<.0001, η_p^2=.83), confirming better performance for distinct nonword pairs, and learning over trials. The main effect of materials failed to reach significance (*F*(1,21) = 0.17, *MSe* = 16.011, *p*=.69). The phonological similarity x trials interaction attained significance (*F*(4,84) = 3.46, *MSe* = 1.032, *p*=.012, η_p^2=.14), demonstrating different rates of learning the similar and distinct nonword pairs. All remaining interactions failed to reach significance (all *ps*>.05). This confirmed parallel patterns of results across materials A and B. A simple main effects analysis yielded significantly better learning of distinct compared with similar nonword pairs at Trials 2 to 5 (all *ps*<.01). Performance was equivalent for the two types of nonword pairs at Trial 1 (*F*(1,22) = 2.39, *MSe* = 0.33, *p*=.14). Figure 3.2(b) shows these results when collapsing over materials A and B.

However, it could be argued that the significant phonological similarity x trials interaction reported for nonword pairs is being driven by near-floor performance for both similar and distinct nonword pairs at Trial 1. In order to determine whether this was the case, a second three-way mixed-design ANOVA was conducted on the nonword data excluding performance at Trial 1. This second ANOVA incorporated two within-subjects factors: phonological similarity (similar nonwords, distinct nonwords) and trials (2 to 5); and one between-subjects factor: materials (materials A, materials B).

The ANOVA revealed significant main effects of phonological similarity ($F(1,21) = 14.44$, $MSe = 5.715$, $p=.001$, $r=.64$) and trials ($F(3,63) = 79.10$, $MSe = 0.911$, $p<.0001$, $\eta_p^2=.79$), confirming better performance for distinct nonword pairs, and learning over trials. The main effect of materials failed to reach significance ($F(1,21) = 0.76$, $MSe = 18.588$, $p=.76$) as did the phonological similarity x trials interaction ($F(3,63) = 1.23$, $MSe = 0.947$, $p=.31$). All remaining interactions also failed to attain significance (all *p*s>.05). On the basis of these results, the significant phonological similarity x trials interaction from the initial analysis appears to be caused by floor performance at Trial 1. In order to confirm this latter finding, an alternative method of assessing rates of learning similar and distinct nonword pairs over trials was conducted. This method involved generating learning gradients (or slopes) based on performance over Trials 2 to 5 for individual participants and comparing these gradients across similar and distinct nonword pairs (e.g. Cumming et al., 2003, 2006; Page et al., 2006). Such an analysis revealed a larger learning gradient for distinct nonword pairs ($M = 1.1$, $SD = 0.5$) compared with similar nonword pairs ($M = 0.8$, $SD = 0.6$), suggesting faster learning of distinct nonword pairs over Trials 2 to 5. A one-tailed paired-samples t-test indicated that this difference approached significance ($t(22) = -1.43, p=.08$).

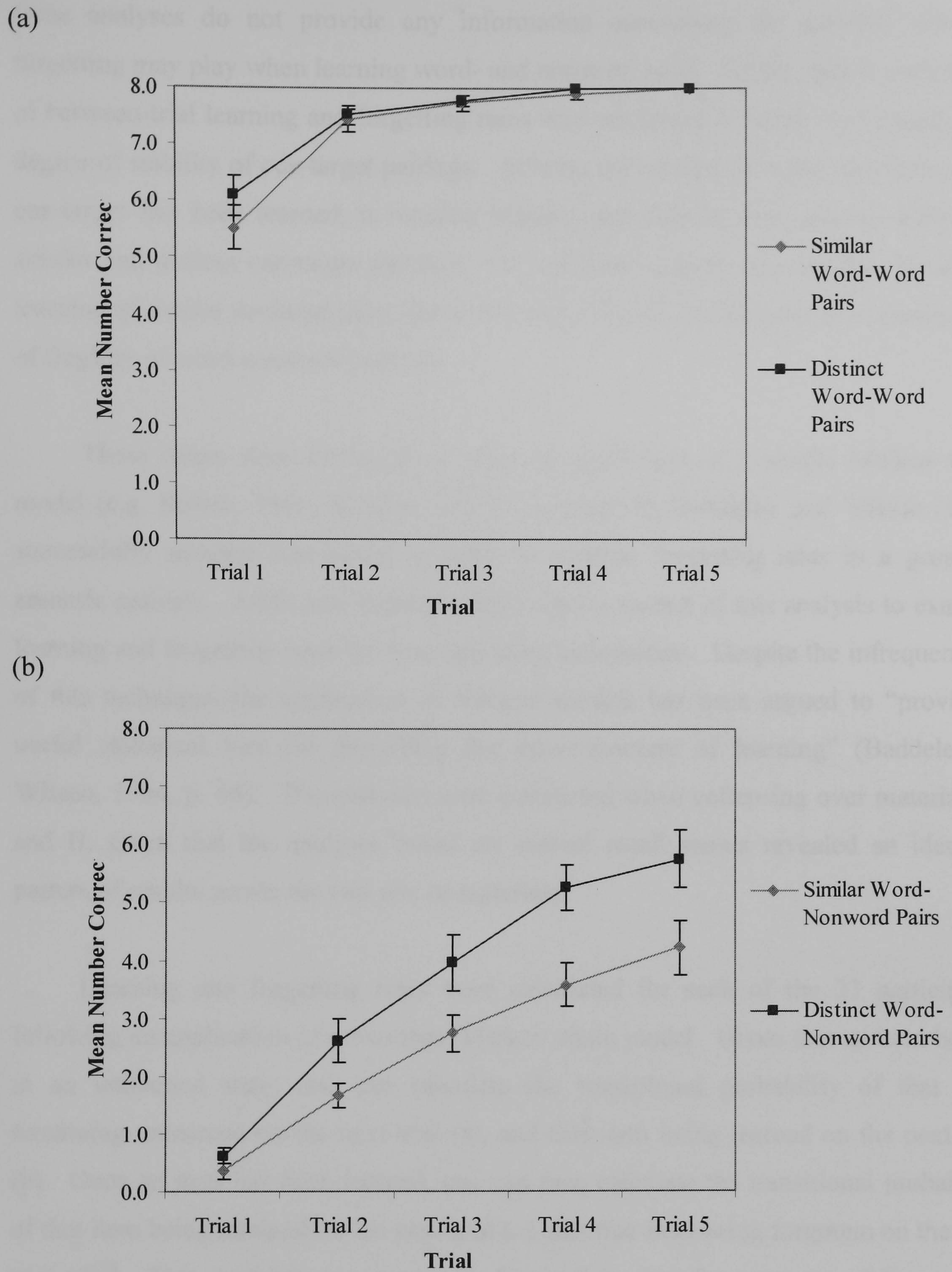


Figure 3.2: *Performance on the paired-associate task as a function of phonological similarity and trials for (a) word-word pairs; and (b) word-nonword pairs (error bars represent standard error of the mean)*

3.3.2.2 Markov Model Analysis

The analyses based on correct recall performance examine the effects of phonological similarity on the rate of learning word- and nonword pairs. However, these analyses do not provide any information concerning the possible role that forgetting may play when learning word- and nonword pairs. To this end, a comparison of between-trial learning and forgetting rates was conducted in order to investigate the degree of stability of cue-target pairings. At issue are whether it is the case that once a cue-target has been learned, it remains learned; and whether this pattern differs for similar and distinct cue-target pairings. Of particular interest was whether the slower learning of similar nonword pairs shown for correct recall performance is a consequence of fragility of word-nonword pairings.

These issues were investigated using an application of a simple Markov chain model (e.g. Bower, 1961; Kintsch, 1977). A study by Baddeley and Wilson (1994) successfully adopted this model in order to analyse forgetting rates in a group of amnesic patients. Addis and Kahana (2004) used a variant of this analysis to examine learning and forgetting rates for item and order information. Despite the infrequent use of this technique, the application of Markov models has been argued to "provide a useful statistical tool for describing the microstructure of learning" (Baddeley & Wilson, 1994, p. 56). The analyses were conducted when collapsing over materials A and B, given that the analyses based on correct recall scores revealed an identical pattern of results across the two sets of materials.

Learning and forgetting rates were calculated for each of the 23 participants following an application of a two-state Markov chain model. Given that an item begins in an unlearned state, one can calculate the transitional probability of that item remaining unlearned on the next trial (a), and that item being learned on the next trial (b). Once an item has been learned, one can then calculate the transitional probability of that item being retained on the next trial (c), and that item being forgotten on the next trial (d)[18]. This model is represented in Figure 3.3. For the purposes of the current analyses, the transitional probabilities (b) and (d) represent learning and forgetting rates, respectively. See Appendix 2(c) for the formulae used to calculate these transitional probabilities.

---

[18] The sum of the transitional probabilities (a) and (b), or (c) and (d), must equal 1.0 given that an item cannot be both unlearned and learned, or retained and forgotten, respectively.

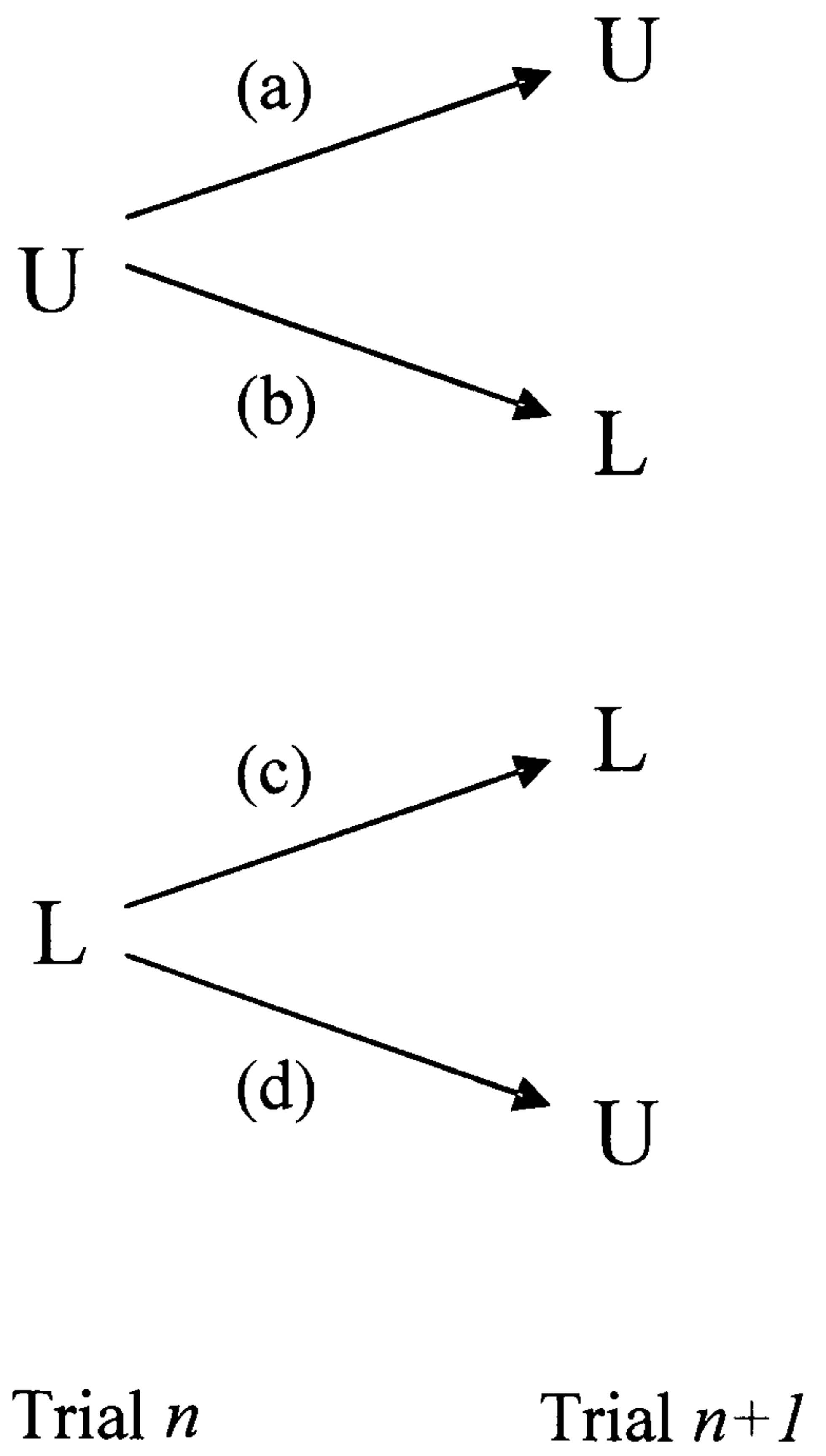


Figure 3.3: *Representation of a two-state Markov chain model. U represents an item in an unlearned state; L represents an item in a learned state. Transitional probabilities across trials are represented as follows: an unlearned item remaining unlearned (a); an unlearned item being learned (b); a learned item remaining learned (c); and a learned item being forgotten (d).*

For each participant, transitional probabilities were calculated for each of four transitional steps: Trial 1 to Trial 2 (T1-T2), Trial 2 to Trial 3 (T2-T3), Trial 3 to Trial 4 (T3-T4), and Trial 4 to Trial 5 (T4-T5), for each of the four experimental conditions. Calculation of transitional probabilities took into account all eight cue-target pairs at each transitional step. Due to the nature of some participants' responses, instances arose where transitional probabilities could not be calculated for (b) and/or (d) at some (or all) of the transitional steps. Failure to obtain (b) reflected situations in which all cue-target pairs were learned on one trial, and remained learned on the next trial; situations in which (d) could not be calculated reflected instances in which no cue-target pairs were learned on one trial, with these cue-target pairs remaining unlearned on the next trial[19]. As a result, participants were excluded if (b) and/or (d) could not be calculated for one (or more) of the transitional steps. These criteria were applied separately to the word and nonword conditions.

---

[19] In both cases the denominator value would be zero rendering any calculation impossible.

Due to ceiling performance in similar and distinct word pair conditions from Trial 3 onwards, 83% (n=19) of participants were excluded. It was therefore considered inappropriate to conduct an analysis of learning and forgetting rates based on the data from only four participants; thus, this data is not reported

For the nonword data, the transitional probability (d) could not be calculated at the transitional step T1-T2 for 19 of the 23 participants. To avoid the elimination of these participants this transitional step was excluded from the analysis. Taking the remaining three transitional steps (T2-T3, T3-T4, T4-T5) into account, 39% (n=9) of participants were excluded. The transitional probability (b) and/or (d) could not be obtained for nine participants. Transitional probability values (b) and (d) were therefore calculated for the remaining 14 participants. Table 3.3 shows the mean learning and forgetting rates for similar and distinct nonword pairs at each transitional step.

Table 3.3: *Mean (and standard deviation) transitional probabilities for learning and forgetting rates at each transitional step for similar and distinct word-nonword pair conditions*

		Transitional Step		
Condition	Rate	T2 – T3	T3 – T4	T4 - T5
Similar Nonwords	Learning	.277 (.187)	.351 (.206)	.439 (.297)
	Forgetting	.185 (.303)	.210 (.293)	.285 (.364)
Distinct Nonwords	Learning	.380 (.247)	.424 (.317)	.620 (.366)
	Forgetting	.054 (.145)	.054 (.100)	.014 (.053)

The pattern of results shows that learning rates increase to a similar degree over transitional steps for both types of nonword pair. Higher learning rates compared to forgetting rates were observed for both similar and distinct nonword pairs. However, whereas higher learning rates were observed for distinct compared with similar nonword pairs at each transitional step, higher forgetting rates were shown for similar compared to distinct nonword pairs at each transitional step. Furthermore, whilst the rate of forgetting increases over transitional steps for similar nonword pairs, the rate of forgetting for distinct pairs is at floor over transitional steps. However, Table 3.3 indicates large degrees of variation in both learning and forgetting rates at each transitional step. Due to this large variation, the data were collapsed over transitional steps for the purposes of statistical analysis. Table 3.4 shows the revised mean learning and forgetting probability values for the two nonword conditions

Table 3.4: *Mean (and standard deviation) transitional probabilities for learning and forgetting rates collapsed over transitional steps for similar and distinct word-nonword pair conditions*

	Condition	
Transition Type	Similar Nonword Pairs	Distinct Nonword Pairs
Learning	.356 (.176)	.475 (.264)
Forgetting	.226 (.241)	.041 (.052)

Participants' transitional probability values were entered into a two-way repeated measures ANOVA incorporating two within-subject factors: transition type (learning, forgetting) and phonological similarity (similar, distinct). The ANOVA revealed a significant main effect of transition type ($F(1,13) = 20.36$, $MSe = 0.054$, $p=0.001$, $r=.78$), confirming higher learning rates compared with forgetting rates. The main effect of phonological similarity failed to reach significance ($F(1,13) = 0.37$, $MSe = 0.041$, $p=.55$), but there was a significant transition type x phonological similarity interaction ($F(1,13) = 6.94$, $MSe = 0.047$, $p=.021$, $r=.59$), reflecting a different pattern of learning and forgetting rates for similar and distinct nonword pairs. A simple main effects analysis yielded a significantly higher forgetting rate for similar nonword pairs in comparison to distinct nonword pairs ($F(1,13) = 8.26$, $MSe = 0.03$, $p=.013$, $r=.62$). Statistically comparable learning rates were reported for the two types of nonword pairs ($F(1,13) = 1.69$, $MSe = 0.06$, $p=.22$).

3.3.2.3 Error Analyses

Error analyses were conducted in order to investigate the degree of stability of the target items. Of issue was whether it is the case that once a target item has been learned, it remains learned; and whether similar and distinct target items show the same pattern. Of particular interest was whether the slower learning observed for similar nonword pairs based on correct recall performance is due to fragility of nonwords.

Participants' responses were scored for a number of pre-defined error types: *omissions, association errors* and *item errors*. Omissions were defined as instances when participants failed to provide a response following the presentation of a cue word. Association errors were defined as instances when a target item was recalled correctly, but was associated to the incorrect cue word (e.g. 'powder' as a response to the cue word 'butter', but where this is actually the correct response to the cue word 'shadow'). Item errors were defined as responses which differed from any of the target items presented within a trial; these were divided into a number of sub-types: *semantically related errors, repetitions of a cue word,* and *phonological errors*. Semantically related errors represented responses that were semantically related to the target item (e.g. 'theatre' as a response for the target word 'concert'). Repetitions of a cue word represented responses that were repetitions of a cue word. Phonological errors represented responses which were phonologically incorrect (e.g. 'dapoth' as a response for the target nonword 'dapeth').

For each participant, the proportions of each error type were calculated for each of the five trials for each of the four experimental conditions. As each trial contained eight cue-target pairs, proportional values were obtained by dividing the number of each type of error by eight. The sum of these four proportional values should equal 1.0 at each trial. As with the correct recall analysis, the word and nonword errors were analysed separately.

Word-Word Pairs: The pattern of errors observed for word pairs showed extremely low proportions of each error type, with a fairly similar pattern of errors across similar and distinct word pairs; see Appendix 2(d). Omissions occurred most frequently (mean of 6% and 5% for similar and distinct word pairs, respectively). Item errors were made infrequently and all represented lexical words (means of <3% for similar and distinct word pairs). Of these item errors, the majority represented semantically related errors (44.4% and 50.0% for similar and distinct word pairs, respectively) and repetitions of a cue word (38.9% and 37.5% for similar and distinct word pairs, respectively). Only a small proportion of item errors were phonological errors (16.7% and 12.5% for similar and distinct word pairs, respectively). Association errors were extremely rare (mean of <3% for both types of word pairs), with these tending to occur at earlier trials. The frequency of omissions and item errors decreased over trials.

Due to the extremely low frequency of association errors and item errors, analyses were conducted on omissions only. Participants' error scores, expressed as proportions, were entered into a two-way repeated measures ANOVA incorporating two within-subjects factors: phonological similarity (similar, distinct) and trials (1 to 5). The ANOVA revealed a significant main effect of trials ($F(1.284,28.256) = 29.73$, $MSe = 0.042$, $p<.0001$, $\eta_p^2 = .58$), indicating a reduction in omissions over trials. The main effect of phonological similarity failed to reach significance ($F(1,22) = 1.24$, $MSe = 0.007$, $p=.28$) as did the phonological similarity x trials interaction ($F(1.684,37.039) = 2.43$, $MSe = 0.008$, $p=.11$); see Figure 3.4

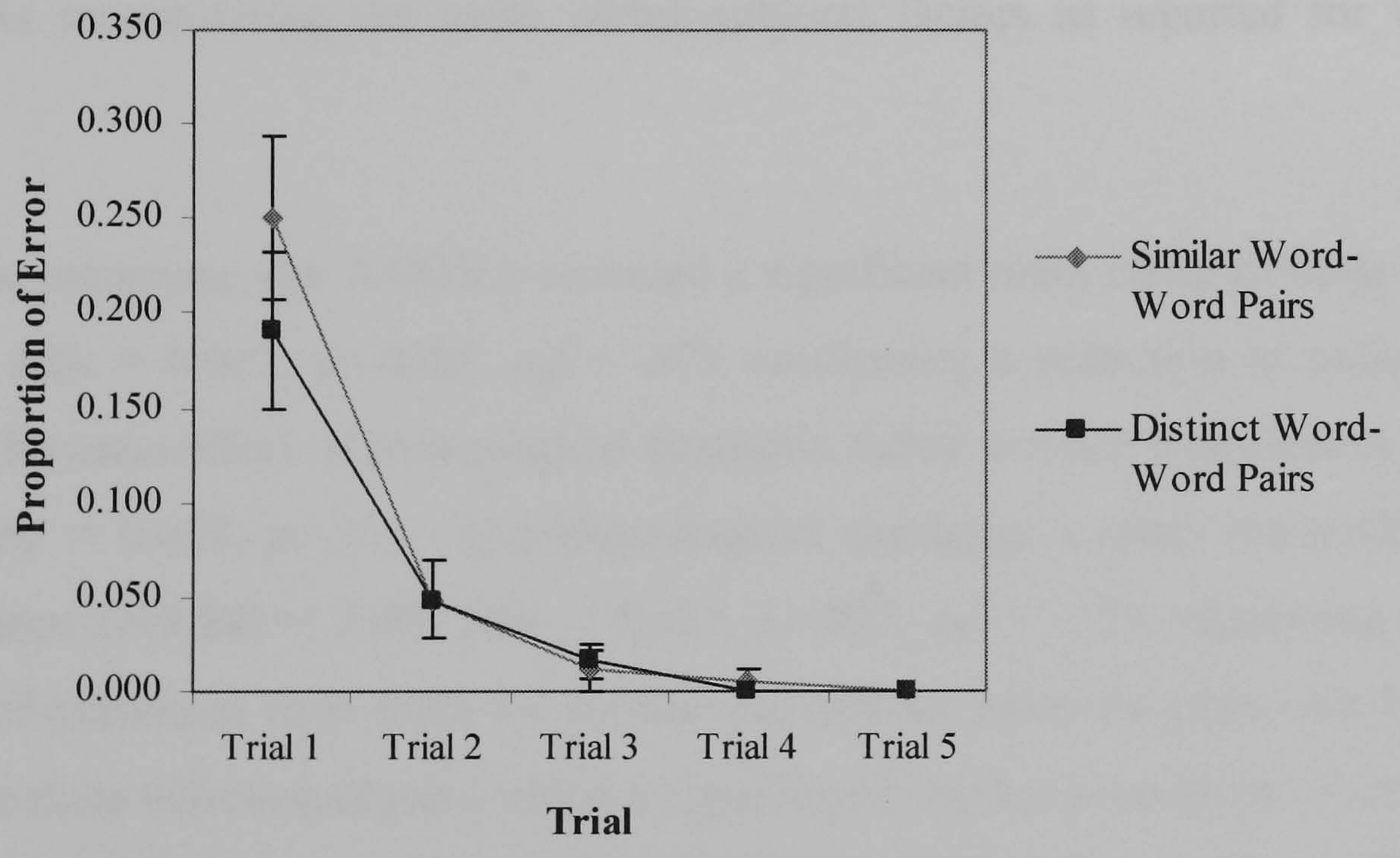


Figure 3.4: *Proportion of omissions as a function of phonological similarity and trials for similar and distinct word-word pairs (error bars represent standard error of the mean)*

Word-Nonword Pairs: The pattern of errors observed for nonword pairs differed from that reported for word pairs. The proportion of errors produced was much higher for nonword pairs. Furthermore, a different pattern of errors emerged for the similar and distinct nonword pairs; see Appendix 2(e). Omissions were made most frequently for distinct nonword pairs (means of 30% for omissions and 23% for item errors), whilst omissions and item errors were made equally frequently for similar nonword pairs (means of 33% and 35%, respectively). The frequency of omissions decreased over trials for both nonword pairs. All item errors were non-lexical in nature and therefore all represented the phonological errors sub-type. No semantically related errors or repetitions of a cue word were made. Phonological errors decreased over trials for distinct nonword pairs, but remained fairly constant over trials for similar nonword pairs. Few association errors were made (mean of <2% for both types of nonword pair), with these occurring at each trial and to a similar degree across both types of nonword pair. Large variation in the frequency of these error types was observed.

Due to the extremely low frequency of association errors, analyses were conducted on omissions and item errors (i.e. phonological errors) only. Participants' scores, expressed as proportions, were entered into separate two-way repeated measures ANOVAs incorporating the same within-subjects factors as reported for word pairs above.

For omissions, the ANOVA revealed a significant main effect of trials ($F(4,88) = 149.25$, $MSe = 0.013$, $p<.0001$, $\eta_p^2 = .87$), confirming a reduction in omissions over trials. The main effect of phonological similarity failed to reach significance ($F(1,22) = 1.10$, $MSe = 0.033$, $p=.31$). The phonological similarity x trials interaction reached significance ($F(4,88) = 2.90$, $MSe = 0.013$, $p=.027$, $\eta_p^2 = .12$), suggesting a different pattern of omissions over trials for similar and distinct nonword pairs (see Figure 3.5). A simple main effects analysis yielded a significantly higher proportion of omissions for similar nonword pairs at Trial 5 only ($F(1,22) = 12.50$, $MSe = 0.01$, p=.002, $r=.60$); equivalent proportions of these errors were found for the two types of nonword pair at Trials 1 to 4 (all *p*s>.05).

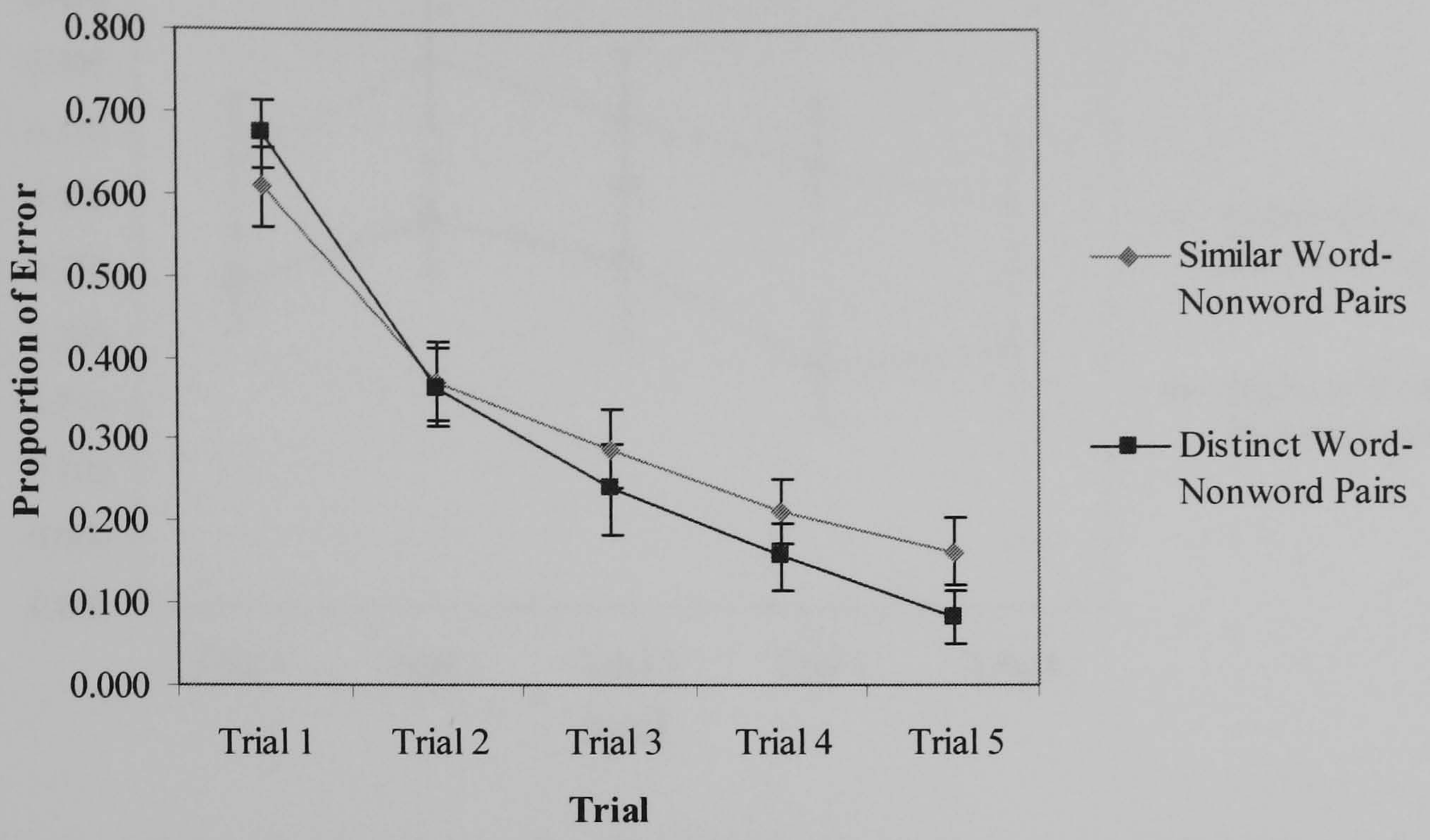


Figure 3.5: *Proportion of omissions as a function of phonological similarity and trials for similar and distinct word-nonword pairs (error bars represent standard error of the mean)*

For phonological errors, the ANOVA revealed a significant main effect of phonological similarity ($F(1,22) = 19.26$, $MSe = 0.039$, $p<.0001$, $r=.68$), demonstrating a higher frequency of phonological errors for similar nonword pairs. The main effect of trials was significant ($F(4,88) = 3.31$, $MSe = 0.025$, $p=.014$, $\eta_p^2 = .13$), suggesting a reduction in these errors over trials. However, polynomial trend analysis reported a non-significant linear trend ($F(1,22) = 3.08$, $MSe = 0.044$, $p=.09$), indicating that phonological errors do not decrease linearly over trials (see Figure 3.6). The phonological similarity x trials interaction failed to reach significance ($F(4,88) = 0.21$, $MSe = 0.026$, $p=.93$).

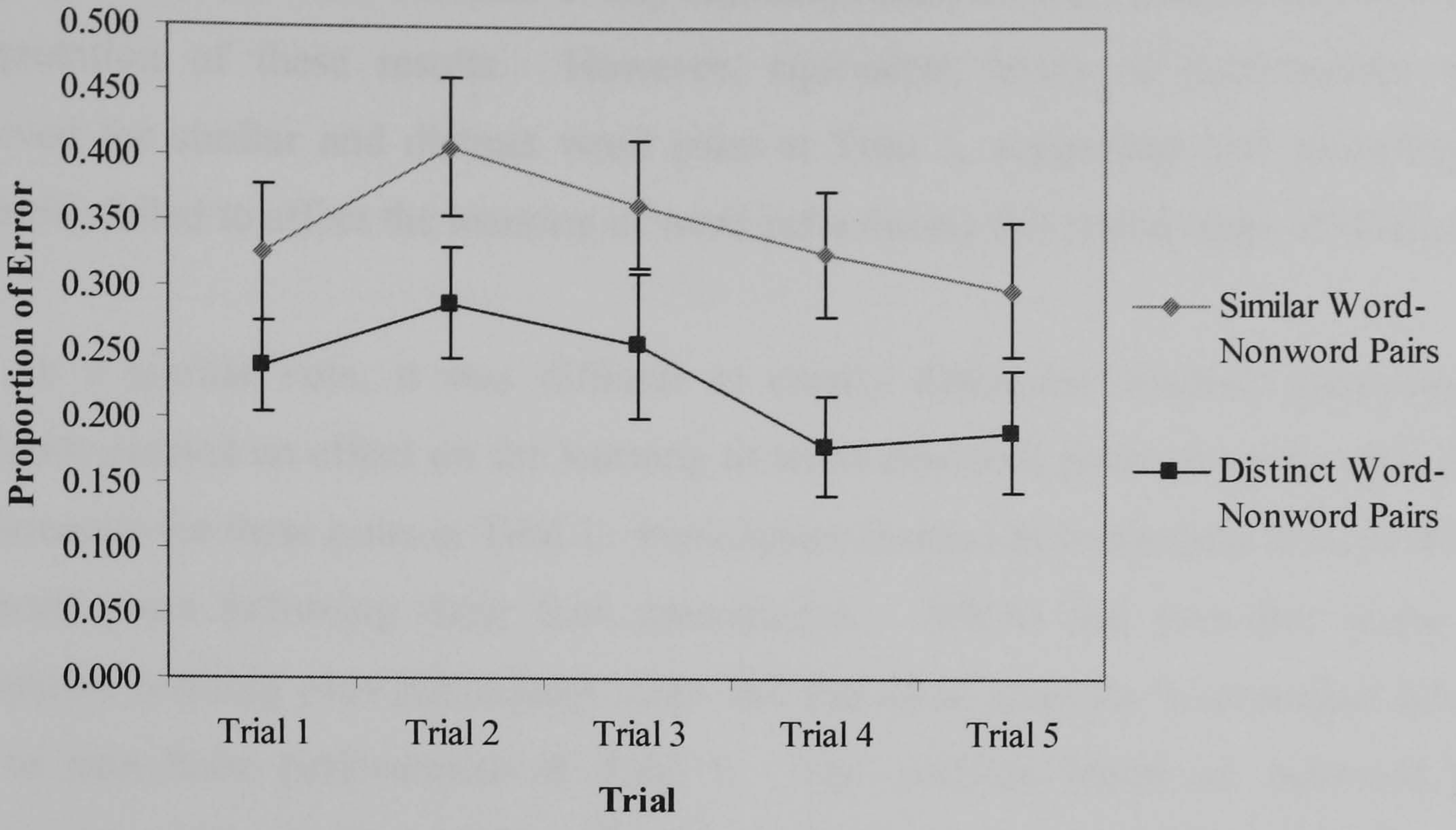


Figure 3.6: *Proportion of phonological errors as a function of phonological similarity and trials for similar and distinct word-nonword pairs (error bars represent standard error of the mean)*

3.3.3 Discussion

The aim of Experiment 2 was to provide a replication and extension of Papagno and Vallar's (1992) experiment in which the effect of phonological similarity on the long-term learning of word-word and word-nonword pairs was investigated. Papagno and Vallar's (1992) design was modified slightly for the purposes of the current experiment. Firstly, English materials were selected. Secondly, the type of nonword used was explicitly controlled; only nonwords subjectively rated as being unwordlike were selected in an attempt to reduce the opportunity for participants to utilise existing lexical knowledge when learning nonword pairs. In line with Papagno and Vallar's (1992) findings, it was predicted that phonological similarity would selectively disrupt the learning of word-nonword pairs, but not the learning of word-word pairs.

The analysis of correct recall performance provides some support for this prediction. The effect of phonological similarity on the learning of word-word pairs could not be reliably assessed due to near-ceiling performance from Trial 2 onwards. Participants learned between 69% and 76% of word pairs following only a single presentation of these pairs; by Trial 3 performance was almost 100%. This high level of performance restricts the outcome of any statistical analysis and subsequently limits the interpretation of these results. However, equivalent levels of performance were observed for similar and distinct word pairs at Trial 1, suggesting that phonological similarity failed to affect the learning of word pairs during this initial stage of learning.

In a similar vein, it was difficult to clearly determine whether phonological similarity exerted an effect on the learning of word-nonword pairs given the near-floor performance for these pairs at Trial 1. Participants learned between only 4% and 8% of nonword pairs following their first presentation. Whilst this provides scope for cumulative learning over subsequent trials, the statistical analysis is somewhat limited due to near-floor performance at Trial 1. The analysis based on nonword pair performance over all five trials indicated that phonological similarity impaired the learning of word-nonword pairs; faster learning over trials was demonstrated for distinct nonword pairs as evidenced by poorer recall performance for similar nonword pairs at Trials 2 to 5. This pattern of results replicates Papagno and Vallar's (1992) findings and extends their results to English materials and participants. This latter result is a novel finding. However, when excluding nonword pair performance at Trial 1 due to near-floor performance, the effect of phonological similarity on nonword pair learning was eliminated, suggesting that similar and distinct nonword pairs are learned at equivalent rates. However, a further analysis based on learning gradients suggested that the distinct nonword pairs showed a marginally non-significantly greater rate of learning compared to the similar nonword pairs. Taking these analyses together, it is suggested that there is some evidence to suggest that phonological similarity impairs the learning of similar nonword pairs; however, this evidence is somewhat limited due to floor effects.

Although the current findings are somewhat limited given the problems associated with ceiling and floor effects for word- and nonword pair performance, respectively, it is argued that the current findings tentatively suggest that PSTM contributes specifically to the learning of novel phonological representations. It has been shown that a variable known to affect the operation of the phonological loop in a clearly defined way may reveal a corresponding effect on the learning of unfamiliar material. It is suggested that phonological similarity exerts a negative effect on nonword pair learning due to the absence of existing lexical-semantic representations of this material in LTM. As a result, participants are forced into relying predominantly on temporary representations of the phonological structure of these nonwords in order to create stable and permanent representations in LTM. Conversely, the learning of word-word pairs appears to proceed without reliance on the phonological structure of this material at the onset of learning. Instead, the learning of familiar material makes use of non-phonological learning codes, such as semantic codes, and as such by-passes PSTM. Indeed, previous studies have shown that LTM is not affected by phonological coding (e.g. Dale & Baddeley, 1969), but is instead affected by semantic similarity (e.g. Baddeley, 1966b; Baddeley & Dale, 1966), providing evidence that individuals are able to utilise semantic coding in LTM. The current findings therefore provide limited evidence that PSTM mediates the learning of unfamiliar material, thereby providing some support for the PSTM hypothesis.

Interestingly, ceiling and floor effects were not observed in Papagno and Vallar's (1992) experiment. Performance at Trial 1 for word pair learning in this original experiment was reported to be between 29% and 43%, with performance failing to reach ceiling even by Trial 5. It is speculated that a possible reason for the ceiling effects in the current experiment is a result of using cue and/or target words with high frequency and/or concreteness values. The current cue and target words were selected following the same criteria as stated in Papagno & Vallar (1992), by having a written frequency of greater than 10.2 per million. However, these authors failed to provide specific details about the frequency values of individual cue and target words, or the indeed the mean frequency of each of their cue and target words sets. Furthermore, Papagno and Vallar (1992) fail to provide further characteristics of their cue and target words such as their grammatical category (i.e. nouns, verbs, adjectives) or their concreteness values. It is therefore assumed that these characteristics were not controlled for or matched in their experiment.

Further support for the idea that Papagno and Vallar (1992) may have selected cue and/or target words with lower frequency and/or concreteness values than in the current experiment comes from the finding that they found significantly better performance for distinct word pairs at Trial 1. The authors proposed that this reflected participants' use of the short-term phonological store during the initial learning phase, suggesting that participants may only begin to develop semantic associations to learn the word-word pairs following the first presentation of these pairs. The finding that phonological similarity had no effect on the learning of similar and distinct word pairs at Trial 1 in the current experiment may suggest that words with higher frequency and/or concreteness values were selected in the current experiment; this in turn may have facilitated the use of semantic strategies at an earlier stage in learning compared to Papagno and Vallar (1992).

A further possible drawback regarding the design of Papagno and Vallar's (1992) experiment refers to the degree of association within word-word pairs; again, the authors provide no information as to whether this was controlled for. Every effort was made in the current experiment to limit the degree of association within word-word pairs in an attempt to reduce the ease with which participants could generate strategies in order to facilitate their memory for these pairs.

The current experiment demonstrated extremely high levels of performance for word pairs compared to Papagno and Vallar (1992). It is tentatively suggested that these experimental design differences may have contributed to the observed differences in the level of word pair learning across the two experiments.

In a similar vein, Papagno and Vallar (1992) did not report such low levels of nonword pair learning at Trial 1 as was shown in the current experiment. Performance at Trial 1 in the original experiment was between 8% and 15%. Furthermore, higher levels of performance were reported at Trial 5 in the original experiment (between 55% and 88%) compared to the current experiment (between 53% and 71%). The better nonword pair learning observed in the original experiment may be in part due to selecting nonwords that only differed from real words by one letter. It is proposed that these nonwords may be more wordlike compared to the nonwords used within the current experiment, which were all subjectively rated as being unwordlike.

Finally, an analysis of between-trial learning and forgetting rates was undertaken using an application of a Markov model in order to investigate the degree of stability of cue-target pairings. Given the ceiling effects for the learning of word pairs, this analysis was conducted on nonword pair learning only. The results suggest that similar and distinct nonword pairs are learned at comparable rates. Interestingly, however, similar nonword pairs showed significantly higher rates of forgetting than distinct nonword pairs. Thus, whilst similar nonword pairs are apparently no more difficult to learn than distinct nonword pairs, they are more unstable and more susceptible to forgetting. In contrast, once a distinct nonword pair has been acquired, it is highly likely to remain intact, given the extremely low forgetting rate observed for this type of nonword pair.

It is important to acknowledge that the analysis based on between-trial learning and forgetting rates using a Markov model and the analysis based on correct recall performance results represent two different ways of examining the data. Learning based on correct recall performance tracks the number of nonword pairs correctly learned on each trial, but does not take into account whether the same nonword pairs are recalled correctly from trial to trial. Analysis of between-trial learning and forgetting rates takes both these points into account. However, that these two analyses represent different ways of investigating the data does not preclude the possibility that the results of one analysis can inform the results of the other analysis.

With this in mind, it is tentatively suggested that the equivalent between-trial learning rates observed for similar and distinct nonword pairs in the Markov model analysis may actually reflect the fact that a larger number of the same distinct nonword pairs are being correctly recalled between trials compared to similar nonword pairs. This would mean that there would be fewer additional distinct nonword pairs to learn on subsequent trials; as a result, the transitional probability that an unlearned distinct nonword pair would be learned on the next trial will increase given that the number of distinct nonword pairs left to be learned decreases between trials. In contrast, the transitional probability that an unlearned similar nonword pair will be learned on the next trial is more constant over trials, given that different similar nonword pairs are learned on individual trials. Tables 3.5 and 3.6 present a hypothetical example of this explanation. This shows that even though there is faster learning of distinct nonword pairs compared with similar nonword pairs based on correct recall performance (see Table 3.5 and Figure 3.7), the between-trial learning rates based on a Markov model analysis are comparable (see Tables 3.6 and 3.7). Importantly, this example serves to illustrate how the two different methods of analysis are related.

Table 3.5: *Data from a hypothetical participant for similar and distinct nonword pairs (√ = correct responses; x = incorrect responses)*

Nonword Pair	Similar Nonword Pairs					Distinct Nonword Pairs				
	Trial 1	Trial 2	Trial 3	Trial 4	Trial 5	Trial 1	Trial 2	Trial 3	Trial 4	Trial 5
1	√	x	√	x	x	√	√	x	√	√
2	√	√	√	x	√	√	√	√	√	√
3	√	x	√	√	√	√	√	√	√	√
4	x	√	√	√	√	x	√	√	√	√
5	x	√	x	x	√	x	x	√	√	√
6	x	x	x	x	√	x	x	x	√	√
7	x	x	x	√	x	x	x	x	x	√
8	x	x	x	x	x	x	x	x	x	x
Number Correct	3	3	4	3	5	3	4	4	6	7

Table 3.6: *Between-trial learning and forgetting rates using a Markov model based on the hypothetical participant's data shown in Table 3.5*

Transitional Probability	Similar Nonword Pairs					Distinct Nonword Pairs				
	T1-T2	T2-T3	T3-T4	T4-T5	*Mean*	T1-T2	T2-T3	T3-T4	T4-T5	*Mean*
Learning Rate	.40	.40	.25	.60	.41	.20	.25	.50	.50	.36
Forgetting Rate	.67	.33	.50	.0	.38	.0	.25	.0	.0	.06

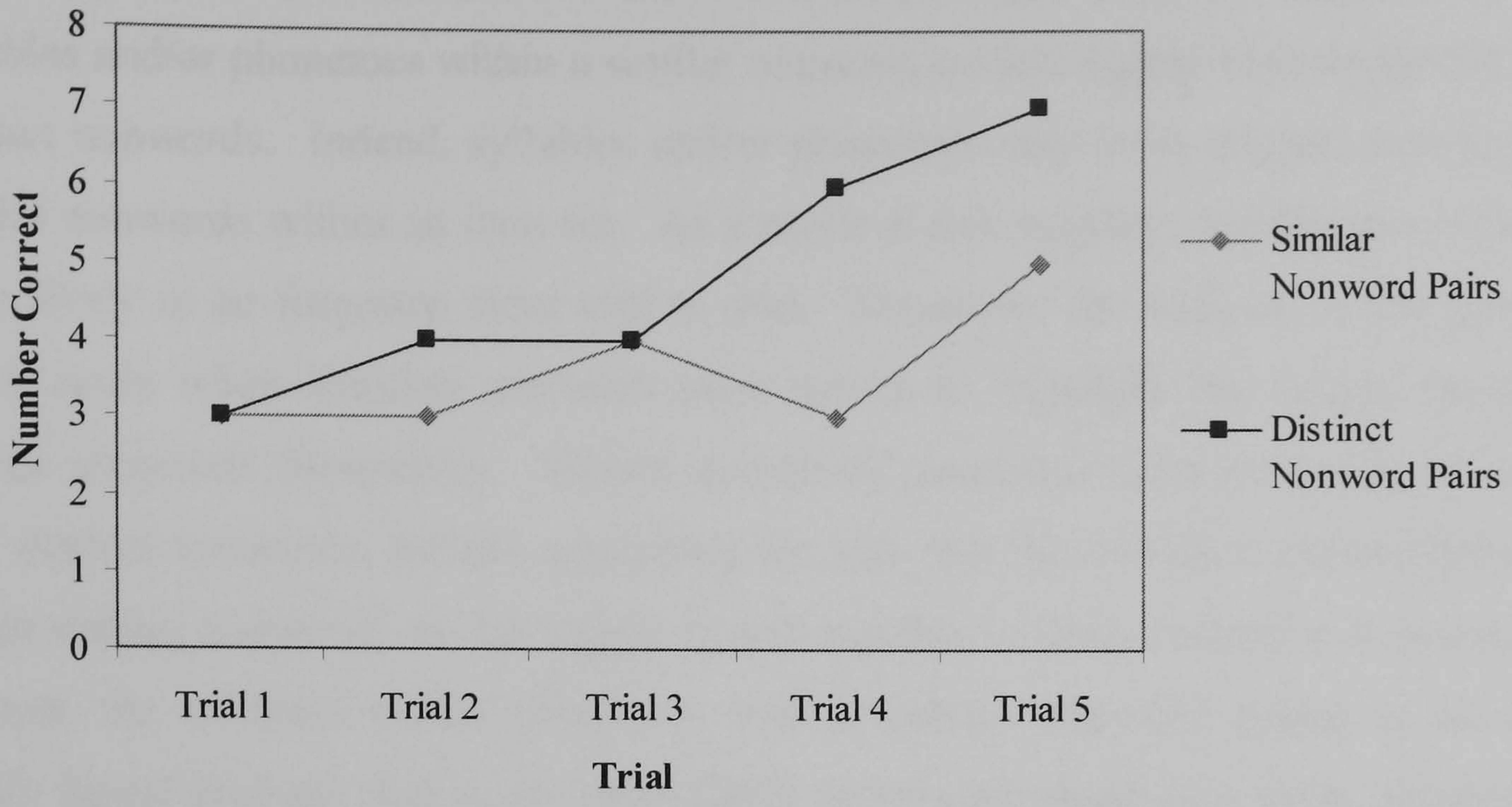


Figure 3.7: *Data from a hypothetical participant showing performance as a function of phonological similarity and trials for word-nonword pairs*

Table 3.7: *Mean transitional probabilities for learning and forgetting transition types for similar and distinct word-nonword pairs based on the hypothetical participant's data in Figure 3.7*

	Condition	
Transition Type	Similar Nonword Pairs	Distinct Nonword Pairs
Learning	.41	.36
Forgetting	.38	.06

The finding of a higher between-trial forgetting rate for similar nonword pairs is interesting as it suggests that similar nonword pairings are particularly fragile in comparison to distinct nonword pairings. One possible interpretation of this pattern of results is that the pairing of a cue word and a similar nonword is more likely to fall apart over trials compared with the pairing of a cue word and a distinct nonword. However, the pattern of association errors revealed by the error analyses suggests that this may not be the case. The rarity of association errors suggests that similar (and distinct) nonwords are not often paired with the wrong cue word. Taking this finding into account, the results from the Markov model analysis may imply that similar nonword pairs are subject to high levels of forgetting between trials due to the fragile nature of individual similar nonwords, rather than nonword pairings. That is, it may be that the syllables and/or phonemes within a similar nonword are less tightly bound together than distinct nonwords. Indeed, syllables and/or phonemes may even migrate between the similar nonwords within an item set. As a result of this fragility, similar nonwords are more likely to be forgotten from trial to trial. Moreover, the analysis of the types of errors made when learning nonword pairs serves to highlight the fragile nature of similar nonwords themselves. Similar nonwords generated more phonological errors than distinct nonwords, further supporting the idea that the syllables and/or phonemes within similar nonwords are not tightly bound together to form a cohesive nonword. In contrast, the syllables and/or phonemes with a distinct nonword appear to be more tightly bound together and so are more likely to be maintained over trials, presumably leading to better learning of distinct nonword pairs.

The analyses of between-trial learning and forgetting rates have served to highlight the potential role forgetting plays when learning nonword pairings. Moreover, these analyses have permitted a novel insight into how nonword pairings and/or individual nonwords are learned which is not revealed when analysing correct recall performance results alone. It is suggested that future experiments within this thesis may benefit from conducting such an analysis in addition to analyses based on correct recall performance. However, it is important to bear in mind that the present experiment has highlighted a potential drawback to conducting analyses of learning and forgetting rates using a Markov model. Ceiling- and floor effects can lead to a reduction in sample size. Furthermore, sample size can also be reduced due to missing data at one or more transitional steps. Future experiments therefore need to concentrate on overcoming these methodological limitations.

The remaining error analyses were not particularly informative in terms of the pattern of learning observed. Omissions were the only error type to show a reduction in frequency over trials for both word and nonword pairs, with a slower reduction in these errors over trials for similar nonword pairs only. This suggests that participants were more likely to attempt to provide a response to a cue word for distinct nonwords. However, phonological similarity did not differentially affect the overall frequency of these errors. No error type revealed a similar pattern of results to those obtained from the analysis of correct recall performance. This finding suggests that analyses based on different error types may not be sensitive enough to capture the differential effect of phonological similarity on nonword pair learning over trials.

3.4 Chapter Summary

Experiment 1 established a set of words and nonwords appropriate for use in Experiment 2. These words and nonwords generated clear PSEs in an auditory ISR task, thereby confirming that phonological similarity had been adequately manipulated. Furthermore, evidence for a lexicality effect was also found. Experiment 2 provided some evidence that PSTM contributes to the learning of unfamiliar material. It was tentatively shown that phonological similarity selectively impaired the learning of similar nonword pairs. In contrast, phonological similarity failed to impair the learning of similar word pairs when performance was assessed at Trial 1 only. These results offer a partial replication and extension of Papagno and Vallar's (1992) original experiment. Furthermore, the results of the analysis when applying a Markov model suggest that phonological similarity has its negative impact on between-trial forgetting rates, suggesting that similar nonword pairs are particularly fragile. Indeed, the results of error analyses further suggest that similar nonwords are also fragile in nature. However, despite the present findings, the conclusions are somewhat restricted due to ceiling and floor effects for the learning of word- and nonword pairs, respectively. Chapter Four therefore presents a further experiment aimed at addressing these limitations.

Chapter Four: Re-Investigating and Extending the Effects of Phonological Similarity on Paired-Associate Learning

4.1 Experiment 3: Effects of Phonological Similarity on Paired-Associate Learning – A Further Attempt to Replicate Papagno and Vallar (1992)

The main aim of Experiment 3 was to provide a further replication of Papagno and Vallar (1992), addressing the limitations encountered in Experiment 2. Due to ceiling and floor effects, performance for word- and nonword pair learning was not matched at Trial 1 in Experiment 2. In order to avoid these effects, it was decided to match performance at this trial. Various matching procedures were investigated in a series of pilot experiments; the outcome of which resulted in the current experimental design. However, given that this design is considerably different from the design used in Experiment 2 and Papagno and Vallar (1992), it is necessary to briefly describe the numerous modifications imposed in these pilot experiments and their corresponding outcomes.

Pilot 3(i): This involved manipulating the number of cue-target pairs to be learned. Given that Experiment 2 presented participants with eight word- and eight nonword pairs, the number of word pairs presented was increased to 10 and 12 pairs, whilst the number of nonword pairs presented was decreased to 4 and 6 pairs. It was predicted that this would reduce word pair learning, whilst boosting nonword pair learning. All other aspects of the design remained the same as in Experiment 2. However, given that the pilot aimed to match performance at Trial 1, only a single trial was presented. Eight participants took part in this pilot.

Surprisingly, the results did not confirm these predictions. Firstly, word pair performance on the single trial remained just as high as in Experiment 2, despite the presentation of 12 word pairs; participants still learned between 63% and 75% of these pairs. Similarly, nonword pair performance on the single trial remained near floor even when only four pairs were presented; participants only learned between 6% and 13% of these pairs. The manipulation of number of pairs to be learned therefore failed to satisfactorily match word- and nonword pair performance.

Pilot 3(ii): This pilot maintained the modification adopted in Pilot 3(i) whilst making several additional modifications. Firstly, the presentation rate was increased to 4 s per pair rather than 4.75 s per pair. This was achieved by reducing the inter-pair interval from 2 s to 1.25 s. This faster presentation rate equated that used in Papagno and Vallar (1992) and Papagno et al. (1991). Secondly, the duration of the two recall phases (digit recall phase and pair recall phase) were modified. In Experiment 2, participants had 10 s to recall the digit sequence and 7 s to recall the target item of a cue-target pair in response to the presentation of a cue item. The next phase of the trial was initiated at the end of these recall periods, regardless of when participants made their response. In the current pilot, the next phase of a trial was initiated immediately following a participant's response, with a maximum time allowed of 10 s for the digit recall phase and 7 s for the pair recall phase. It was anticipated that this latter revision would provide more control over what activity participants engage in during the recall phases and would reduce the time available to further rehearse, or indeed forget, the cue-target pairs. The final modification addressed the idea that Experiment 2 utilised cue words with high frequency and/or concreteness values, as previously discussed in Chapter Three. With this in mind, the cue words were changed. Following Papagno et al. (1991), the revised cue words each had a concreteness value of <4.0. Given that these revised words represented more abstract words (e.g. *humble* or *restore*) than those used in Experiment 2 (e.g. *baby* or *tractor*), it was anticipated that this would reduce the ease with which semantic associations could be generated within word pairs. It was predicted that, taken together, these revisions would reduce word pair performance and enhance nonword pair performance at Trial 1. A single trial was presented. Eight participants volunteered to take part in this pilot.

The revisions succeeded in lowering word pair performance on the single trial; participants learned between 21% and 23% of 12 word pairs. In contrast, nonword pair performance remained at floor with participants learning between 0% and 6% of four nonword pairs. Thus, despite numerous modifications, this pilot failed to equate word- and nonword pair performance.

Pilot 3(iii): The design of Pilot 3(ii) was repeated but with the instructions given to participants changed. Participants were instructed to attempt to concentrate on learning only a small number of pairs rather than attempting to learn the entire set. It was predicted that this would encourage participants to learn only one or two pairs for both word- and nonword pairs. A single trial was presented. Eight participants participated.

This pilot succeeded in enhancing nonword pair performance on the single trial; participants learned between 15% and 25% of four nonword pairs. However, a corresponding improvement was observed for word pair performance on the single trial; participants learned 53% of 12 word pairs. This pilot therefore failed to match word- and nonword pair performance.

Pilot 3(iv): The next pilot adopted the minimal paired-associate learning task (e.g. Baddeley & Warrington, 1970; Peterson, 1966). The primary aim of this task was to match word- and nonword pair performance based on mean word- and nonword pair spans. These spans would then be used to investigate learning over trials in Experiment 3 by presenting participants with mean word- and nonword pair spans plus 100%. For example, if mean word pair span was 5.0 and mean nonword pair span was 2.5 in the minimal paired-associate task, then 10 word pairs and 5 nonword pairs would be presented at Trial 1 in Experiment 3. It was anticipated this procedure would match word- and nonword pair performance at Trial 1 in Experiment 3.

In the minimal paired-associate task, participants were initially presented with a small number of cue-target pairs to learn on a single trial. The number of these pairs was then increased by one until a maximum number of pairs had been reached. For word pairs, participants began with the presentation of three pairs; this increased by one pair up to a maximum of seven word pairs. For nonword pairs, participants began with the presentation of one pair; this was increased by one pair up to a maximum of four pairs. At each pair length, participants were presented with two different sets to learn. For example, for word pairs, participants would be presented with one set containing three word pairs, followed by a second set containing a different three word pairs. The same procedure would be followed at each pair length. Average word- and nonword spans were identified as the mean number of cue-target pairs learned. Twenty participants volunteered to take part in this task[20].

---

[20] This larger sample size was due to assigning lexicality as a between-subjects factor in this experiment.

The results revealed an unexpected pattern. Mean word pair span was calculated as 2.6 (out of a maximum of 7). This was considerably lower than expected based on mean word pair performance in Pilots 3(i) and 3(iii); performance in these pilots reached 8.3 and 6.4 words (out of a maximum of 12), respectively. Applying the span plus 100% criteria based on the current mean word pair span would mean presenting participants with seven word pairs at Trial 1 in Experiment 3. This was considered inappropriate given that participants are capable of learning approximately this number of word pairs on a single trial, as evidenced in Pilots 3(i) and 3(ii). Mean nonword span was calculated as 0.2 (out of a maximum of 4). This was also lower than expected given that nonword pair performance reached 0.4 and 0.8 (out of a maximum of 4) for Pilots 3(i) and 3(iii), respectively. Indeed, it would not even be feasible to apply the span plus 100% criteria at Trial 1 in Experiment 3 based on a nonword span of 0.2 nonwords.

Given the failure of Pilots 3(i) to 3(iv) to match word- and nonword pair performance on a single trial, it was decided to abandon the idea of matching word- and nonword pair performance at Trial 1 in Experiment 3. Instead, an alternative and novel design was adopted. The design followed that adopted in Pilot 3(ii) but with two modifications. Firstly, six word- and nonword pairs were presented. Secondly, whereas word pairs were presented over five trials (as in Experiment 2), nonword pairs were presented over an extended period of up to 12 trials. The idea was to select the trial at which nonword pair performance matched that of Trial 1 for word pair performance. This nonword pair trial would then be taken to reflect 'revised Trial 1', with the subsequent four trials reflecting 'revised Trials 2 to 5'. Performance on these five trials would then be compared against the five word pair trials. In this sense, nonword pair performance would match Trial 1 word pair performance.

A secondary aim of Experiment 3 was to further investigate the learning of nonword pairs. Experiment 2 controlled for degree of wordlikeness of nonwords; in contrast, this variable was manipulated in the current experiment. Chapter One reviewed research supporting the idea that long-term lexical knowledge influences PSTM performance (e.g. Hulme et al., 1991, 1995, 1997; Gathercole et al., 1991a, 1999; Gathercole, 1995; Greg et al., 1989; Roodenrys et al., 1994, 2002; Roodenrys & Hinton, 2002; Thorn & Frankish, 2005; Thorn et al., 2005). Furthermore, a number of studies have provided evidence that knowledge of the phonological structure of language can influence PSTM. For example, nonwords rated as wordlike have been shown to be better recalled compared to nonwords rated as unwordlike (e.g. Gathercole et al., 1991a, 1999; Gathercole, 1995a; Grant et al., 1997; Vitevitch et al., 1997; Roodenrys & Hinton, 2002; Thorn & Frankish, 2005; Thorn et al., 2005; von Bon et al., 1997).

If it is the case that PSTM for wordlike nonwords benefits from the existing structure of language to a greater extent than unwordlike nonwords, it may be predicted that this benefit for wordlike nonwords would lead to differences in the long-term learning of these nonwords compared to the learning of unwordlike nonwords. Indeed, there is some support for this hypothesis. Gathercole et al. (1996, cited in Gathercole & Martin, 1996) conducted a study in which adults learned word-nonword pairs differing in their degree of wordlikeness. The wordlike nonword pairs were learned significantly faster than the unwordlike nonword pairs. Given this finding, it is of particular theoretical interest to the current thesis to further investigate the potential influence of existing knowledge of the structure of language on the learning of nonword pairs. Thus, the extent to which phonological similarity may impair the learning of wordlike nonword pairs was examined. The current experiment therefore compares the learning of unwordlike nonword pairs (henceforth referred to as unwordlike pairs), as examined in Experiment 2, with the learning of wordlike nonword pairs (hereafter referred to as wordlike pairs).

Several predictions were made. Firstly, on the basis of the results from Papagno and Vallar (1992), it was predicted that phonological similarity would selectively disrupt the learning of unwordlike pairs, but not the learning of word pairs. Regarding the inclusion of wordlike pair learning, it was predicted that significantly faster learning of wordlike pairs would be observed in comparison to unwordlike nonword pairs, as previously shown by Gathercole et al. (1996, cited in Gathercole & Martin, 1996). However, no firm predictions were made regarding the effects of phonological similarity on the long-term learning of wordlike pairs. If participants are able to capitalise on the existing lexical and phonological structure of language, phonological similarity may have a small effect on learning wordlike pairs when compared with unwordlike pairs. It may be that the wordlike pair results will more closely parallel those observed for word pairs in Experiment 2 and Papagno and Vallar (1992).

4.1.1 Method

4.1.1.1 Design

Participants undertook a paired-associate learning task as described in Experiment 2 with the exception that they were presented with six cue-target pairs to learn on each trial. Furthermore, these pairs were presented over the course of 5 trials for word-word pairs or up to 12 trials for word-nonword pairs. The experiment used a 2 x 3 x 2 mixed-design involving one within-subject factor: phonological similarity (similar items, distinct items); and two between-subjects factors: lexicality (words, wordlike nonwords, unwordlike nonwords) and materials (materials A, materials B). In addition, a further within-subject factor of trials was incorporated: Trials 1 to 5 for word-word pairs, and Trials 1 to 12 for word-nonword pairs. The lexicality and phonological similarity factors combined to produce six experimental conditions: similar word pairs; distinct word pairs; similar wordlike pairs; distinct wordlike pairs; similar unwordlike pairs; and distinct unwordlike pairs.

Participants were randomly assigned to one of the three lexicality conditions. Within each lexicality condition, the factor of phonological similarity was counterbalanced across participants so that half received the similar cue-target pairs followed by the distinct cue-target pairs, and half received the reverse. Participants within each lexicality condition were randomly divided so that half received materials A, and half received materials B (as in Experiments 1 and 2). The same constraints were applied as in Experiment 2 regarding the random presentation of the cue-target pairs on each trial at both learning and recall phases.

4.1.1.2 Participants

Forty-eight undergraduate and postgraduate students attending the University of York volunteered to participate for either course credit or payment. There were 36 females and 12 males aged between 18 years and 42 years (mean age of 20.5 years).

Participants all spoke English as their native language and reported no known hearing or language impairments.

4.1.1.3 Apparatus

The same apparatus were used as in Experiments 1 and 2.

4.1.1.4 Materials

All cue items were two-syllable words. The target words and unwordlike nonwords were selected from those used in Experiment 2. The target wordlike nonwords were selected from the wordlikeness ratings task conducted prior to Experiment 1. Each set contained six cue-target pairs.

Cue Word Sets

Twelve sets of six cue words were constructed; two sets (materials A and B) for each of the six experimental conditions (e.g. *devil, humble, shallow, retain, talent, predict*). Individual cue words had frequency values of <60 and concreteness values of <4.00. Each set was matched for written frequency per million (Kucera & Francis, 1967; mean range across sets: 19 to 27) and concreteness ratings (mean range across sets: 2.87 to 3.42) using the MRC Psycholinguistic Database (Coltheart, 1981). Two separate one-way between subjects ANOVAs revealed no statistically significant differences between the 12 cue word sets for written frequency ($F(11,60) = 0.20$, $MSe = 233.067$, $p=.99$) or concreteness values ($F(11,60) = 0.63$, $MSe = 1865.603$, $p=.80$). See Appendix 3(a) for the cue word sets.

Target Word Sets

Four sets of six target words were selected from the corresponding 8-item target word sets used in Experiment 2: two phonologically similar sets (materials A and B) and two phonologically distinct sets (materials A and B). Two words from each original 8-item set were excluded. Each of the four sets were matched on mean written frequency per million (Kucera & Francis, 1967; mean range across sets: 30 to 35) and concreteness values (mean range across sets: 4.34 to 4.78) using the MRC Psycholinguistic Database (Coltheart, 1981). A number of two-tailed independent t-tests revealed no statistically significant differences between any of the four word sets for either frequency or concreteness values (all *p*s>.05). See Appendix 3(b) for the target word sets.

Phonologically Similar and Distinct Word Sets: In each phonologically similar set, three different initial consonants and three (materials B) or four (materials A) final consonants were used. In each phonologically distinct set, different initial and final consonants were used for each of the six words.

Target Unwordlike Nonword Sets

Four sets of six target nonwords were selected from the corresponding 8-item target nonword sets used in Experiment 2: two phonologically similar sets (materials A and B) and two phonologically distinct sets (materials A and B). Two nonwords from each original 8-item set were excluded. The four sets were matched on mean wordlikeness ratings (mean range across sets: 1.9 to 2.3). A number of two-tailed independent t-tests revealed no statistically significant differences in wordlikeness ratings between any of the four sets (all *p*s>.05). See Appendix 3(c) for the target unwordlike nonword sets.

Phonologically Similar and Distinct Unwordlike Nonword Sets: In each phonologically similar set, three different initial consonants and four final consonants were used. In each phonologically distinct set, different initial and final consonants were used for each of the six nonwords.

Target Wordlike Nonword Sets

Four sets of six target nonwords were chosen from those nonwords with the highest wordlikeness ratings: two phonologically similar sets (materials A and B) and two phonologically distinct sets (materials A and B); see Table 4.1. Each of the sets was constructed following the same criteria as were adopted for the construction of the phonologically similar and distinct unwordlike nonword sets. The four sets were matched on mean wordlikeness ratings (mean range across sets: 2.6 to 3.0). A number of two-tailed independent t-tests revealed no statistically significant differences in wordlikeness ratings between any of the four sets (all *ps*>.05). See also Appendix 3(d) for the target wordlike nonword sets and their properties.

Table 4.1: *Phonologically Similar and Distinct Wordlike Nonword Sets*

Phonologically Similar Wordlike Nonword Sets		Phonologically Distinct Wordlike Nonword Sets	
Materials A	Materials B	Materials A	Materials B
meppict	lappint	putchel	webbist
mellib	lattip	fabbor	mordast
pedmin	parvit	turlict	purldam
peflin	pattish	sappesh	higgart
sellict	rasbit	bergops	roskurl
sempib	raftip	darpist	vernash

Phonologically Similar Wordlike Nonword Sets: Each of the six nonwords had a similar first vowel and a similar second vowel (e.g. *mellib*, *pedmin*); the vowels used within each set were the same as those used in the unwordlike nonword sets. Three different initial consonants and three (materials A) or four (materials B) different final consonants were used.

Phonologically Distinct Wordlike Nonword Sets: Each nonword had a different ordering of vowels for the first and second vowel positions (e.g. *sappesh*, *higgart*). Different initial and final consonants were used for each of the six nonwords within a set.

Target Unwordlike Sets vs. Target Wordlike Sets

Wordlikeness ratings were compared across target unwordlike and wordlike sets to ensure these represented two distinct nonword types. Since no significant differences were found between the four nonword sets for each type of nonword, these were combined. Mean (SD) wordlikeness ratings were 2.1 (0.4) and 2.8 (0.4) for unwordlike and wordlike nonwords, respectively. A two-tailed independent t-test revealed a statistically significant difference between the two types of nonword ($t(46) = -5.88$, $p<.0001$), confirming lower wordlikeness ratings for unwordlike nonwords.

Construction of Cue-Target Pairs

The cue words and target items were paired together following the same constraints as described in Experiment 2. See Appendix 3(e) for the cue-target pair sets.

Editing of Speech Signals

Each cue word and wordlike nonword was recorded onto a minidisc player in a soundproof room by the same speaker as used in Experiments 1 and 2. Each speech signal was edited as described in Experiment 1. The edited speech signals of the target words and unwordlike nonwords from Experiment 2 were re-used.

4.1.1.5 Procedure

The paired-associate task was conducted with individual participants in a quiet testing room. Experimental instructions differed slightly depending on which lexicality condition participants were assigned. Participants allocated to the word conditions were informed the experiment would involve learning sets of six word-word pairs and would take approximately 25-30 mins. Participants assigned to the nonword conditions were advised the experiment involved learning sets of six word-nonword pairs and would take approximately 45-60 mins. All participants were advised that each trial would be divided into three phases: a listening phase; a digit recall phase; and a pair recall phase. The instructions regarding each phase were the same as in Experiment 2 with the exception that participants were allowed a maximum recall period of 10 s to recall the digit sequence, and 7 s to recall the second member of a pair. Participants were informed that these three phases made up a single trial. Participants learning word pairs were advised that five trials would make up one set, those participants learning nonword pairs were instructed that 12 trials would make up one set. Instructions regarding the order in which the pairs would be presented in the listening and pair recall phases, the use of beeps throughout the experiment, the importance of clearly articulating responses, and the recording of participants' responses on to a minidisc player were as in Experiment 2.

Two practice trials, each containing six pairs (not in the experimental test sets) were given prior to the experimental trials.

The procedure followed that of Experiment 2 but with two modifications. The first was an increase in cue-target presentation rate. The second involved initiating the next phase of a trial immediately following participants' responses in the digit recall and pair recall phases. Each pair was presented at a rate of 2.75 s with a 1.25 s inter-pair interval. The ends of the digit recall and pair recall phases were signalled by beeps immediately following participants' responses, unless the maximum recall time period for that phase was reached before a response was provided. Participants completing the word conditions were presented with all five trials. Participants completing the nonword conditions were presented with all 12 trials with one constraint – testing was terminated if all six word-nonword pairs were correctly recalled on four consecutive trials.

Participants' responses were scored as in Experiment 2. The maximum score possible was 6 for each trial, leading to a total score of 30 for word conditions and 72 for nonword conditions. For those participants in the nonword conditions whose testing was terminated early, the maximum score of 6 was automatically assigned to the remainder of the untested trials.

4.1.2 Results

4.1.2.1 Correct Recall Analysis

Responses were scored following the same strict procedure as described in Experiment 2. Instances in which a participant consistently made the same incorrect mispronunciation in response to the same correct cue word were scored as correct[21]; this

---

[21] The same strict criteria were followed as in Experiment 2, with the exception that consistently mispronounced items were scored as correct if they occurred on six or more consecutive trials, which had to include the final trial (i.e. on at least Trials 7 to 12).

occurred on three occasions[22]. Instances in which the correct pronunciation of an item could not be clearly determined were scored as correct; this occurred on three occasions[23].

None of the data sets were normally distributed: word conditions were negatively skewed at later trials; nonword conditions were positively skewed at earlier trials and negatively skewed at later trials. This skewness reflects near-ceiling (word and nonword data) and near-floor (nonword data) performance. Thus, a cautionary note needs to accompany the outcome of all statistical analyses and their subsequent interpretation. No outliers were detected[24].

The means and standard deviations were firstly calculated based on the 8 participants receiving materials A and the 8 participants receiving materials B, for each of the three lexicality conditions separately. In order to avoid conducting a complex four-way ANOVA for the main statistical analysis, the data from each lexicality condition were firstly entered into separate three-way mixed-design ANOVAs to confirm parallel patterns of results across materials A and B. Each of these analyses revealed a parallel pattern of results across the two sets of materials. Table 4.2 therefore shows the means and standard deviations based on the 16 participants in each of the lexicality conditions when collapsing over materials A and B.

Table 4.2 shows that learning was observed over trials in all three lexicality conditions. However, learning was restricted due to near-ceiling performance for similar and distinct word pairs from Trial 3 onwards. Conversely, performance was near floor at Trial 1 for similar and distinct wordlike and unwordlike pairs, thereby permitting the opportunity for substantial learning over remaining trials. The results also demonstrated better learning for word pairs compared to both types of nonword pairs, irrespective of their phonological similarity. Furthermore, better learning was observed for wordlike pairs than unwordlike pairs. However, differential effects of phonological similarity emerged for the word and nonword data. Phonological

---

[22] The wordlike nonword '*putchel*' was mispronounced as '*puckchel*' on two occasions; the unwordlike nonword '*tafflost*' was mispronounced as '*taffloss*' on one occasion.

[23] The wordlike nonword '*putchel*' could not be clearly determined on two occasions; the unwordlike nonword '*berpict*' could not be clearly determined on one occasion.

[24] Outliers were defined and calculated as in Experiment 2.

similarity did not appear to differentially affect word pair learning; the similar and distinct word pairs were learned to comparable degrees at Trials 2 to 5, with better performance for distinct word pairs at Trial 1 only. In contrast, phonological similarity differentially affected the learning of both wordlike and unwordlike pairs. Distinct wordlike pairs showed better learning at all 12 trials compared with similar wordlike pairs. For the unwordlike conditions, whilst distinct unwordlike pairs showed better performance than similar unwordlike pairs at Trials 4 to 12, similar unwordlike pairs revealed better learning at Trials 1 to 3.

Table 4.2: *Mean scores (and standard deviations) obtained from the paired-associate task for each lexicality condition collapsed over materials A and B (maximum score of 6)*

	Trial											
Condition	1	2	3	4	5	6	7	8	9	10	11	12
Similar Word Pairs	2.3 (1.3)	4.9 (1.1)	5.4 (1.1)	5.5 (0.9)	5.7 (0.5)	-	-	-	-	-	-	-
Distinct Word Pairs	3.1 (1.6)	4.6 (1.4)	5.4 (0.9)	5.6 (1.0)	5.8 (0.5)	-	-	-	-	-	-	-
Similar Wordlike Nonword Pairs	0.5 (0.6)	1.3 (1.1)	2.2 (1.6)	2.5 (1.8)	2.9 (1.9)	3.4 (1.9)	3.6 (2.0)	4.2 (2.0)	4.4 (1.8)	4.6 (1.9)	4.6 (1.6)	4.9 (1.6)
Distinct Wordlike Nonword Pairs	0.6 (0.8)	1.9 (1.5)	3.4 (1.5)	4.0 (1.6)	4.4 (1.7)	4.7 (1.4)	4.8 (1.4)	5.0 (1.5)	5.3 (1.4)	5.1 (1.3)	5.2 (1.6)	5.3 (1.3)
Similar Unwordlike Nonword Pairs	0.4 (0.6)	1.1 (1.0)	1.8 (1.2)	1.6 (1.2)	2.3 (1.7)	2.3 (1.3)	2.6 (1.7)	3.4 (1.5)	3.5 (1.8)	4.3 (1.6)	4.5 (1.7)	4.8 (1.2)
Distinct Unwordlike Nonword Pairs	0.1 (0.3)	0.8 (1.1)	1.4 (1.3)	2.2 (1.7)	2.8 (2.3)	3.7 (2.1)	4.5 (2.1)	4.6 (2.0)	4.6 (2.1)	4.7 (2.2)	4.9 (1.7)	4.9 (1.6)

As one of the aims of the current experiment was to select a nonword pair trial at which performance matched that of Trial 1 for word pair performance, the nonword trial which most closely matched word pair performance at Trial 1 was selected as revised Trial 1, with the subsequent four trials selected as revised Trials 2 to 5. This matching methodology was applied to each of the four nonword conditions separately. To this end, Trial 3 was selected as revised Trial 1 for both similar and distinct wordlike pairs, with Trials 4 to 7 as revised Trials 2 to 5. For similar and distinct unwordlike pairs, Trial 5 was selected as revised Trial 1, with Trials 6 to 9 as revised Trials 2 to 5. These data are shown in Figure 4.1, which confirms that word- and nonword pair performance is reasonably well matched at Trial N (where N=1 for word pairs, N=3 for wordlike pairs, and N=5 for unwordlike pairs). As expected, word pair performance exceeds nonword pair performance as learning proceeds over trials. However, performance for the wordlike and unwordlike pairs is comparable at each of the revised trials.

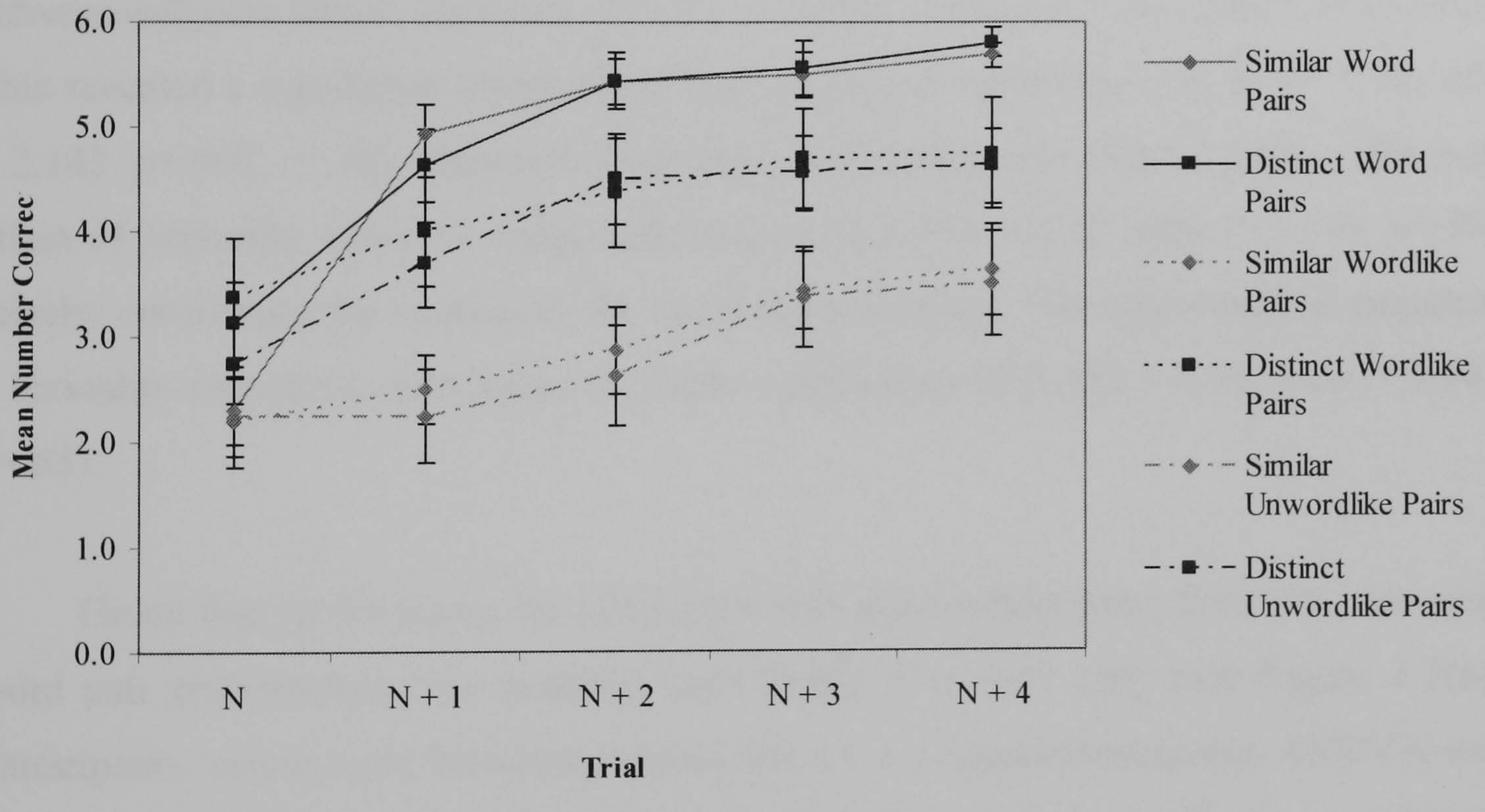


Figure 4.1: *Performance on the paired-associate task as a function of phonological similarity, lexicality and trials when nonword pair performance matches that of Trial N for word pair performance. N represents trial number with N=1 for word pairs; N=3 for wordlike pairs; and N=5 for unwordlike pairs (error bars represent standard error of the mean).*

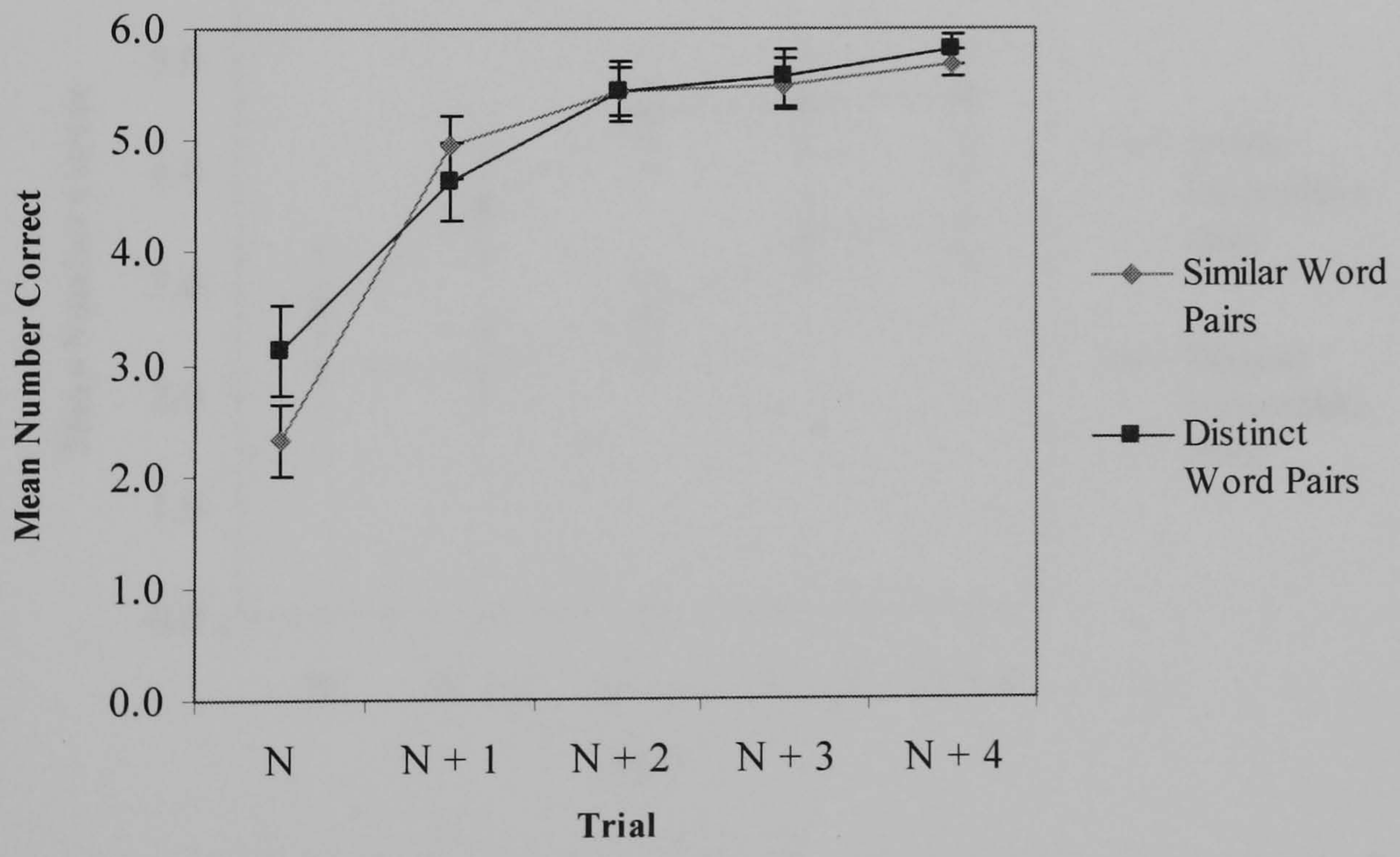
Participants' scores were entered into a three-way mixed-design ANOVA incorporating two within-subjects factors: phonological similarity (similar items, distinct items) and trials (N to N+4); and one between-subjects factor: lexicality (words, wordlike nonwords, unwordlike nonwords). The ANOVA violated the assumption of homogeneity of variance, as indicated by significant Levene's tests: variances were significantly different for similar conditions at Trials N+2, N+3 and N+4 (all *ps*<.05) and distinct conditions at all five trials (all *ps*<.05). These significant differences are most likely being driven by the near-ceiling performance for the word data. It was therefore considered inappropriate to report the statistical results of this ANOVA given the violation of the assumption of homogeneity of variances. Instead, it was deemed more appropriate to conduct analyses on the word and nonword data separately.

However, it was first of all necessary to determine whether the word and nonword data were matched at Trial N, given that one of the aims of the current experiment was to select a nonword pair trial at which performance matched that for word pairs at Trial 1. A two-way mixed-design ANOVA was therefore conducted based on word and nonword pair performance at Trial N only. The ANOVA incorporated one within-subjects factor: phonological similarity (similar items, distinct items); and one between-subjects factor: lexicality (words, wordlike nonwords, unwordlike nonwords). This revealed a significant main effect of phonological similarity (*F*(1,45) = 7.78, *MSe* = 2.143, *p*=.008, *r*=.38), demonstrating better performance for distinct pairs. The main effect of lexicality failed to reach significance (*F*(1,45) = 0.22, *MSe* = 3.176, *p*=.80), thereby confirming the success of the matching procedure. The phonological similarity x lexicality interaction also failed to attain significance (*F*(2,45) = 0.44, *MSe* = 2.143, *p*=.65).

Given that performance for word pairs was near-ceiling from Trial N+2 onwards, word pair performance was assessed over Trials N to N+1 only (see Figure 4.2(a)). Participants' scores were therefore entered into a 2 x 2 repeated-measures ANOVA with two within-subjects factors: phonological similarity (similar words, distinct words) and trials (N to N+1). This revealed a non-significant main effect of phonological similarity (*F*(1,15) = 0.55, *MSe* = 1.833, *p*=.47). The main effect of trials reached significance (*F*(1,15) = 137.27, *MSe* = 0.496, *p*<.0001, *r*=.95), demonstrating learning over trials. The phonological similarity x trials interaction also attained significance (*F*(1,15) = 6.64, *MSe* = 0.763, *p*=.021, *r*=.55), suggesting a different pattern of learning over trials for similar and distinct word pairs. A simple main effects analysis yielded significant learning over trials for distinct word pairs (*F*(1,15) = 30.00, *MSe* = 0.60, *p*<.0001, *r*=.82) and similar word pairs (*F*(1,15) = 83.73, *MSe* = 0.66, *p*<.0001, *r*=.92). Thus, this interaction suggests faster learning over trials for similar compared to distinct word pairs. However, comparable levels of performance were observed for similar and distinct word pairs at Trial N (*F*(1,15) = 4.12, *MSe* = 1.28, *p*=.06) and at Trial N+1 (*F*(1,15) = 0.59, *MSe* = 1.31, *p*=.45).

For the nonword data (see Figures 4.2(b) and 4.2(c)), participants' scores were entered into a three-way mixed-design ANOVA incorporating two within-subjects factors: phonological similarity (similar nonwords, distinct nonwords) and trials (N to N+4); and one between-subjects factor: nonword type (wordlike, unwordlike). The ANOVA revealed significant main effects of phonological similarity (*F*(1,30) = 16.95, *MSe* = 7.599, *p*<.0001, *r*=.60) and trials (*F*(2.937,88.119) = 38.97, *MSe* = 0.870, *p*<.0001, η_p^2=.57), demonstrating better performance for distinct pairs and learning over trials. The main effect of nonword type failed to reach significance (*F*(1,30) = 0.12, *MSe* = 18.641, *p*=.73). All interactions failed to reach significance (all *p*s>.05).

(a)



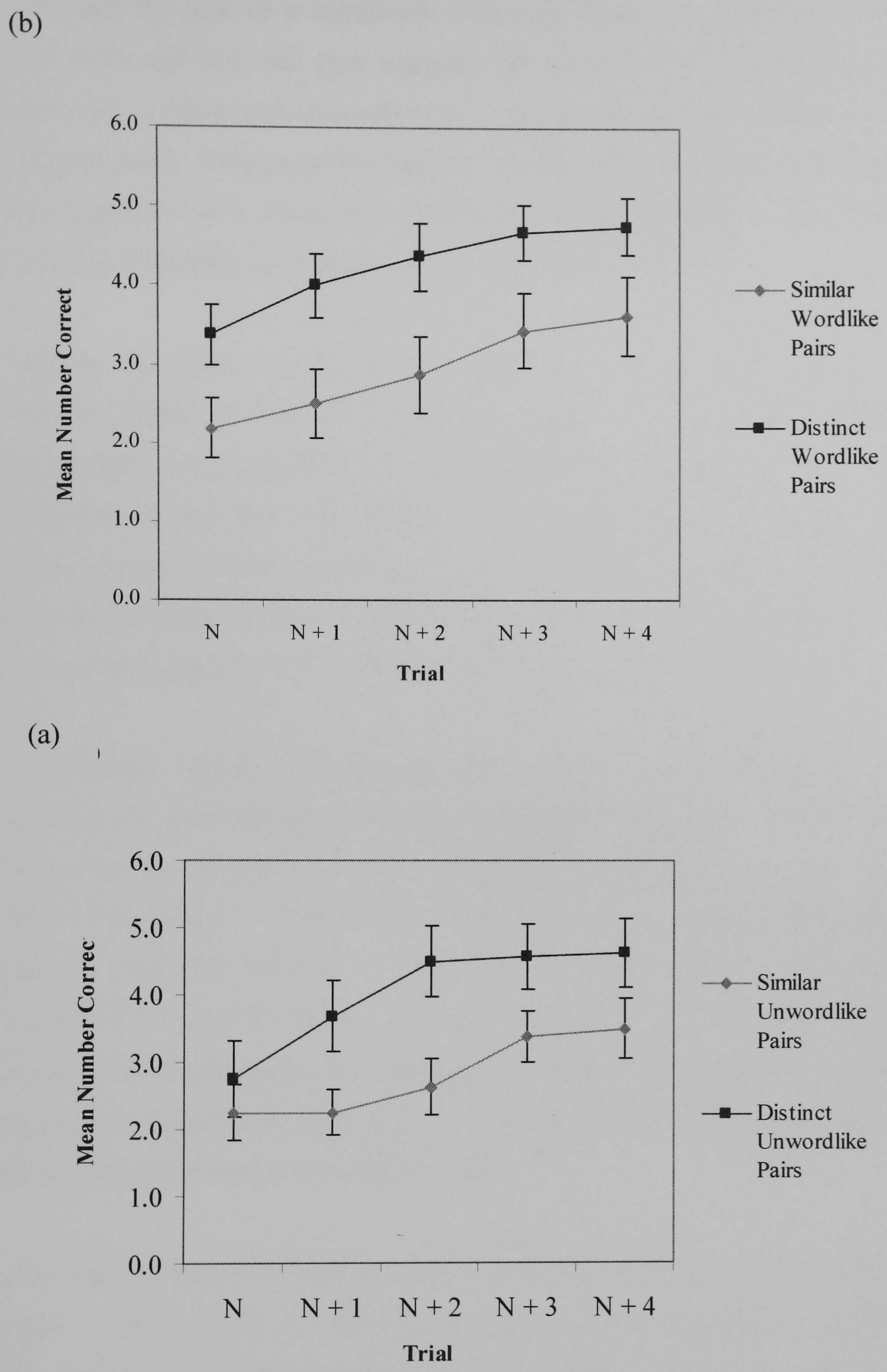


Figure 4.2: *Relationship between phonological similarity and trials for (a) word conditions; (b) wordlike conditions; and (c) unwordlike conditions. N represents trial number with N = 1 for word pairs; N=3 for wordlike pairs; and N=5 for unwordlike pairs (error bars represent standard error of the mean)*

Although the lack of a significant three-way interaction between phonological similarity, trials and nonword type suggests that the pattern of learning over trials is comparable for wordlike and unwordlike pairs, visual comparison of Figures 4.2(b) and 4.2(c) suggest subtle differences between the wordlike and unwordlike conditions, with the latter suggestive of a phonological similarity x trials interaction, with a trend for better learning of distinct unwordlike pairs at Trials N+1 to N+4.

Despite the success in matching performance at Trial N, the analysis on the nonword data failed to find any differential effects of phonological similarity on learning wordlike and unwordlike pairs. Given that the matching procedure resulted in the elimination of more than 50% of trials in the nonword conditions, it was decided to investigate learning in these conditions over all 12 trials in order to determine whether phonological similarity influences nonword pair learning over an extended learning period. A second analysis was therefore conducted on the nonword data only.

Participants' scores were entered into a three-way mixed-design ANOVA incorporating the within-subjects factors: phonological similarity (similar nonwords, distinct nonwords) and trials (1 to 12); and the between-subjects factor: nonword type (wordlike, unwordlike). The ANOVA demonstrated significant main effects of phonological similarity ($F(1,30) = 7.93$, $MSe = 12.321$, $p=.008$, $r=.46$) and trials ($F(3.325.99.752) = 152.27$, $MSe = 2.992$, $p<.0001$, $\eta_p^2=.84$), indicating higher recall performance for distinct pairs, and learning over trials. The main effect of nonword type approached significance ($F(1,30) = 3.11$, $MSe = 29.609$, $p=.09$), suggesting a trend for better recall performance for wordlike pairs.

The trials x nonword type interaction attained significance ($F(3.325,99.752) = 2.96$, $MSe = 2.992$, $p=.031$, $\eta_p^2=.09$), demonstrating a different pattern of learning over trials for each nonword type (see Figure 4.3). A simple main effects analysis yielded significantly better performance for wordlike pairs than unwordlike pairs at each of Trials 2 to 6 (all ps<.05); performance was equivalent at all remaining trials (all ps>.05).

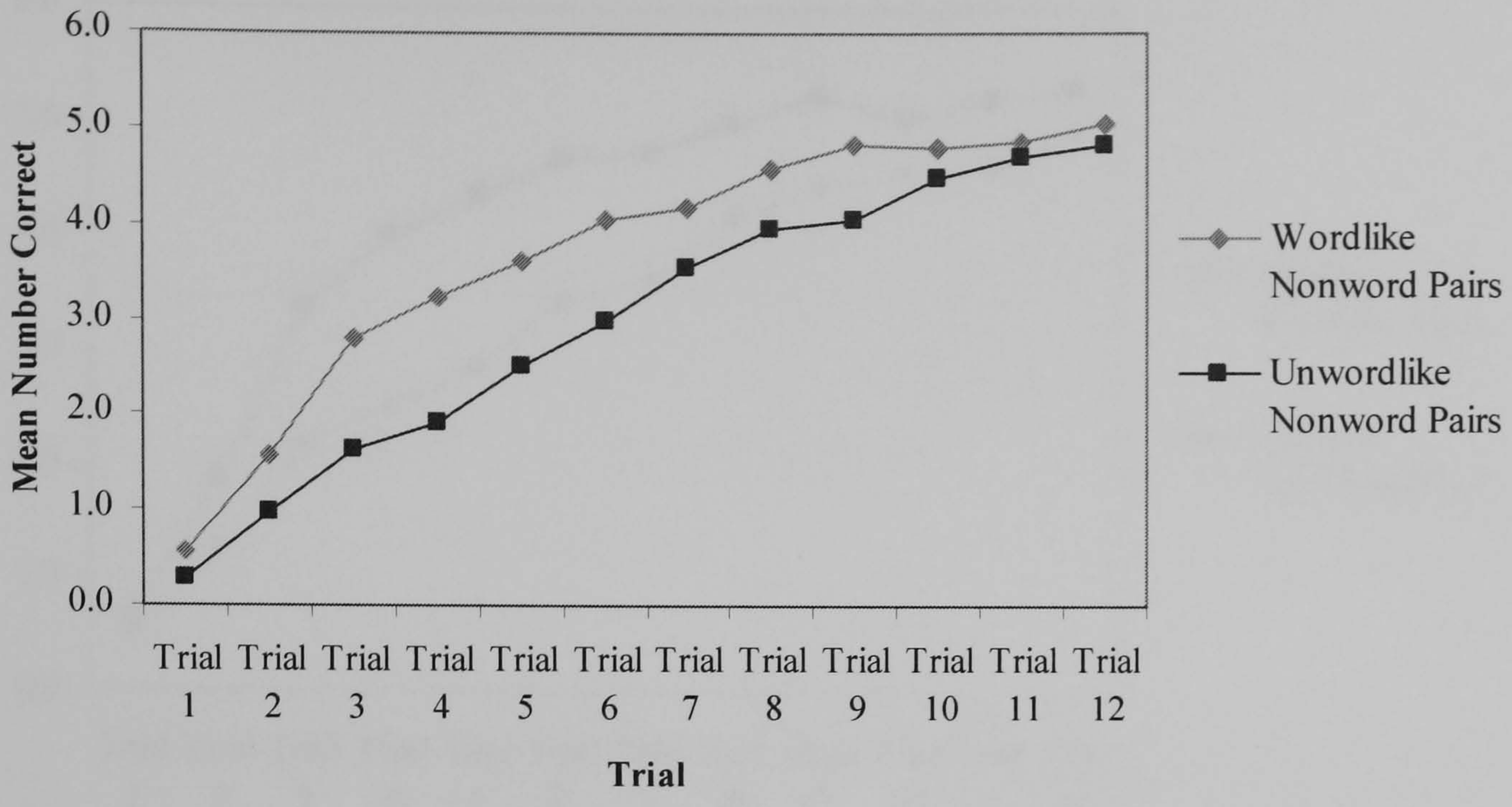


Figure 4.3: *Relationship between nonword type and trials*

The two-way interaction between phonological similarity and nonword type did not reach significance (*F*(1,30) = 0.38, *MSe* = 12.321, *p*=.54). However, a significant phonological similarity x trials interaction emerged (*F*(4.428,132.853) = 4.81, *MSe* = 2.026, *p*=.001, η_p^2=.14), indicating a different pattern of learning over trials for similar and distinct pairs. Moreover, this was incorporated into a significant three-way interaction (*F*(5.463,163.904) = 2.29, *MSe* = 1.642, *p*=.043, η_p^2=.07)[25]. As a means of refining the extent of the effect of phonological similarity on learning over trials, the data from each nonword type were submitted to separate 2 x 12 repeated-measures ANOVAs with phonological similarity and trials as factors (see Figure 4.4).

---

[25] Degrees of freedom were corrected using the Huynh-Feldt estimate. The Greenhouse-Geisser estimate produced a marginally non-significant three-way interaction (F(4.428,132.853) = 2.29, MSe = 2.026, p=.057). Averaging these *p* values obtained *p*=.050, hence the decision to report the Huynh-Feldt estimate (Field, 2005).

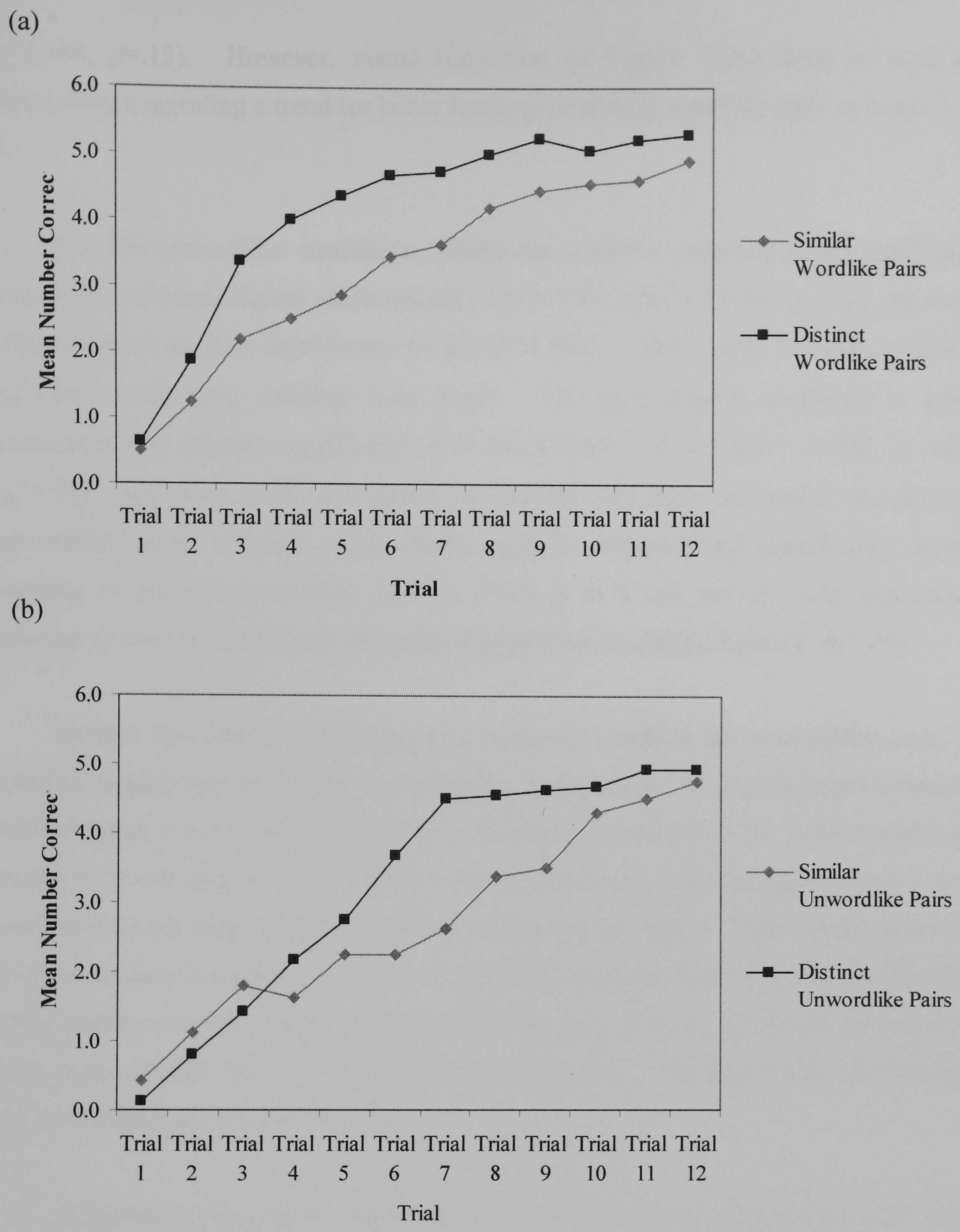


Figure 4.4: *Relationship between phonological similarity and trials for (a) wordlike conditions; and (b) unwordlike conditions*

For the wordlike conditions, the ANOVA revealed significant main effects of phonological similarity ($F(1,15) = 8.33$, $MSe = 8.716$, $p=.011$, $r=.60$) and trials ($F(4.093,61.398) = 77.17$, $MSe = 2.271$, $p<.0001$, $\eta_p^2=.84$), confirming better performance for distinct wordlike pairs and learning over trials. The phonological similarity x trials interaction failed to attain significance ($F(3.698,55.476) = 1.89$, MSe

= 2.484, *p*=.13). However, visual inspection of Figure 4.4(a) hints at such an interaction, suggesting a trend for better learning of distinct wordlike pairs at Trials 3 to 9.

For the unwordlike conditions, whilst the ANOVA reported a non-significant main effect of phonological similarity (*F*(1,15) = 1.87, *MSe* = 15.926, *p*=.19), the main effect of trials attained significance (*F*(2.127,31.903) = 78.01, *MSe* = 4.984, *p*<.0001, η_p^2=.84), confirming learning over trials. The phonological similarity x trials interaction also attained significance (*F*(3.192,47.886) = 5.30, *MSe* = 2.744, *p*=.003, η_p^2=.26), suggesting a different pattern of learning over trials for similar and distinct unwordlike pairs. Simple main effects analysis demonstrated significantly better learning of distinct unwordlike pairs at Trials 6 to 9 (all *p*s<.05), with equivalent learning of similar and distinct unwordlike pairs at all remaining trials (all *p*s>.05).

Despite this pattern of learning over trials for wordlike and unwordlike pairs, it could be argued that the results are somewhat compromised by near-floor performance for both types of nonword pairs at Trial 1. In order to confirm that the present pattern of results were not due to floor effects, a second three-way mixed-design ANOVA was conducted on the nonword data excluding performance at Trial 1. This second ANOVA revealed an identical pattern of results to the initial analysis, thereby confirming that the initial results were not driven by floor performance at Trial 1. All future reference to these nonword data will therefore refer to the results based on the learning of wordlike and unwordlike pairs over all 12 trials.

In summary, the analysis based on matching nonword pair performance with Trial N for word pair performance was compromised due to ceiling effects for the word pairs at Trials N+2 onwards. However, performance was shown to be equivalent at Trial N for the three types of pairs, confirming that nonword pair performance was satisfactorily matched at Trial N for word pair performance. The analysis based on word pairs revealed no effect of phonological similarity on overall recall performance. However, faster learning was observed over Trials N to N+1 for similar word pairs. The results for nonword pairs revealed an identical pattern of results for wordlike and unwordlike pairs. Phonological similarity impaired overall recall performance but failed to disrupt the learning of similar and distinct pairs, as evidenced by comparable learning rates over trials. However, there was a possible trend in the data to suggest that phonological similarity may have slowed down the learning of similar unwordlike pairs.

Analysis of the nonword conditions over all 12 trials revealed a trend for better overall performance for wordlike pairs. Moreover, wordlike pairs showed faster learning over trials. The effect of phonological similarity on wordlike pairs impaired overall performance but failed to disrupt the learning of these pairs, as shown by statistically equivalent learning rates. However, there was a trend in the data to suggest slower learning for similar wordlike pairs. For unwordlike pairs, comparable levels of overall performance were found for similar and distinct pairs. However, phonological similarity slowed down the learning of similar pairs, with poorer performance for these pairs at Trials 6 to 9.

4.1.2.2 Markov Model Analysis

In line with Experiment 2, a comparison of between-trial learning and forgetting rates was attempted using an application of a Markov model in order to further investigate the degree of stability of cue-target pairings.

Learning and forgetting rates were calculated for the 16 participants in each of the three lexicality conditions following the same strict procedure as described in Experiment 2. Learning and forgetting rates were obtained by calculating transitional probabilities as in Experiment 2. However, due to near-ceiling performance (word and nonword data) and near-floor performance (nonword data), too many participants' data had to be omitted for a meaningful analysis of learning and forgetting rates to be conducted.

4.1.2.3 Error Analyses

Participants' responses in each of the three lexicality conditions were scored for omissions, association errors and item errors following the same criteria as in Experiment 2. The proportions of each of these error types were calculated for each trial (5 for word pairs, 12 for nonword pairs). The maximum number of each error type was six; proportional values were therefore obtained by dividing the number of each error type by six. Item errors were further categorised as semantically related errors, repetitions of a cue word, and phonological errors, as described in Experiment 2. Error analyses were analysed separately for word, wordlike and unwordlike pairs.

Word-Word Pairs: The pattern of errors produced showed low proportions of each error type, with a similar pattern across similar and distinct pairs; see Appendix 3(f). Omissions were made most frequently (means of 13% and 11% for similar and distinct pairs, respectively). Few item errors were made and all represented lexical words (means of <6% for both types of word pairs). Of these item errors, the majority represented phonological errors (62.5% and 60.0% for similar and distinct pairs, respectively) and repetitions of a cue word (29.2% and 40.0% for similar and distinct pairs, respectively). Semantically related errors were made for similar pairs only (8.3%). Association errors were rare (means of <5% for both types of word pairs), with these tending to occur at earlier trials. The frequency of omissions and item errors decreased over trials.

Analyses were conducted on omissions only due to the extremely low frequency of association errors and item errors. Participants' error scores, expressed as proportions, were entered into a two-way repeated-measures ANOVA incorporating the factors of phonological similarity (similar, distinct) and trials (1 to 5). The ANOVA (see Figure 4.5) revealed a significant main effect of trials ($F(1.922,28.823) = 36.37$, $MSe = 0.038$, $p<.0001$, $\eta_p^2=.71$), confirming a reduction in omissions over trials. The main effect of phonological similarity failed to reach significance ($F(1,15) = 0.31$, $MSe = 0.020$, $p=.59$) as did the phonological similarity x trials interaction ($F(4,60) = 0.11$, $MSe = 0.010$, $p=.98$).

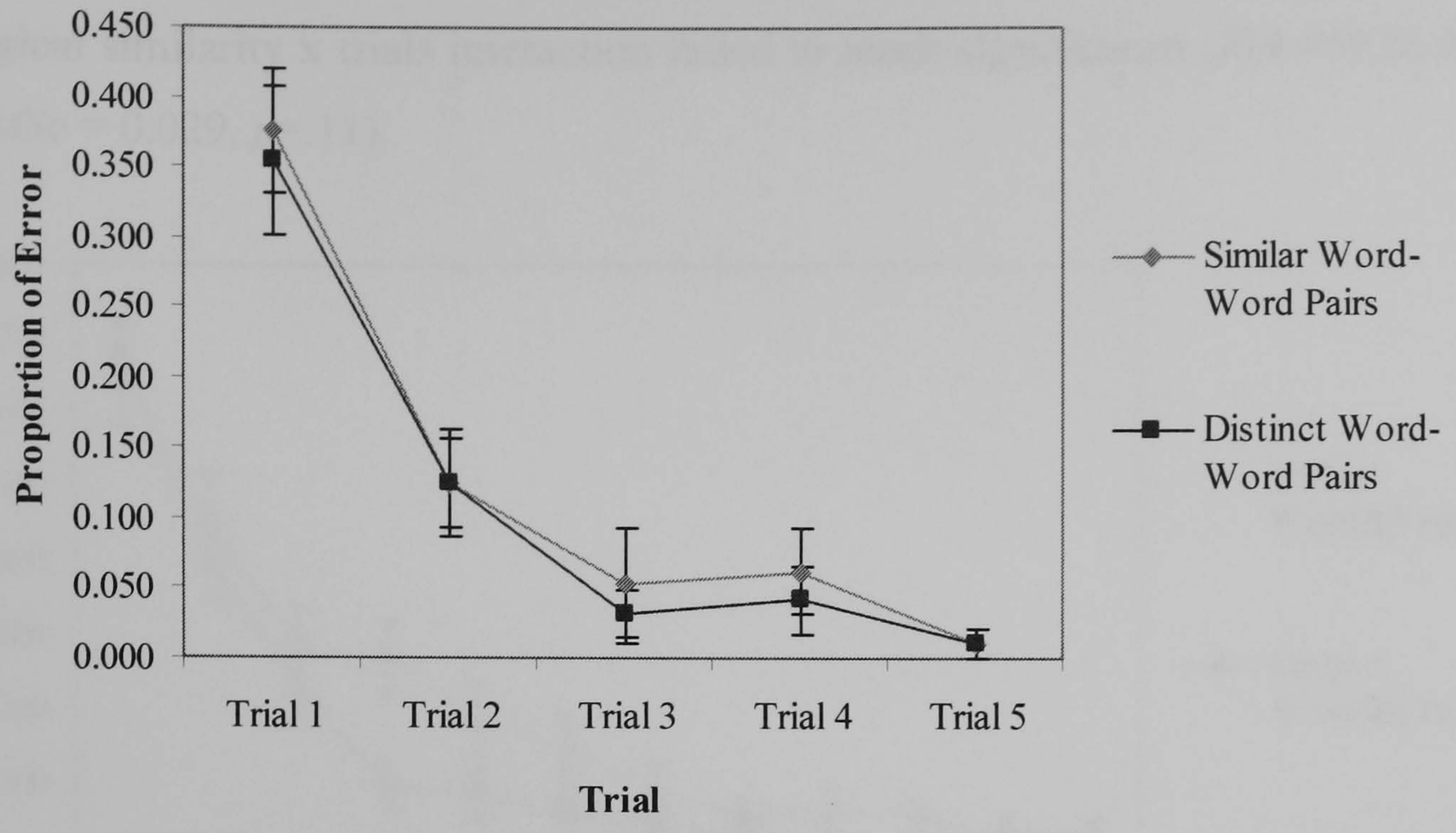


Figure 4.5: *Proportion of omissions as a function of phonological similarity and trials for similar and distinct word pairs (error bars represent standard error of the mean)*

Word-Wordlike Pairs: The pattern of errors revealed comparable proportions of omissions (mean of 14%) and item errors (mean of 16%) for distinct pairs. In contrast, a higher proportion of item errors (mean of 24%) compared to omissions (mean of 19%) were made for similar pairs. The frequency of omissions and item errors decreased over trials for both types of wordlike pairs. The majority of item errors were non-lexical in nature (99%) and as such represented phonological errors. No semantically related errors or repetitions of a cue word were made. Few association errors were made (means of <5% for both types of wordlike pair), with these occurring across trials. See Appendix 3(g).

Error analyses were conducted over all 12 trials. Analyses were conducted on omissions and item errors (i.e. phonological errors) only due to the extremely low frequency of association errors. Participants' error scores, expressed as proportions, were entered into a two-way repeated-measures ANOVA incorporating the same within-subjects factors as reported for word pairs.

For omissions, the ANOVA (see Figure 4.6) revealed significant main effects of phonological similarity ($F(1,15) = 5.09$, $MSe = 0.056$, $p=.039$, $r=.50$) and trials ($F(3.446,51.162) = 79.16$, $MSe = 0.050$, $p<.0001$, $\eta_p^2=.84$), confirming a higher proportion of omissions for similar pairs, and a reduction in these errors over trials. The phonological similarity x trials interaction failed to reach significance (*F*(4.459,66.890) = 1.94, *MSe* = 0.029, *p*=.11).

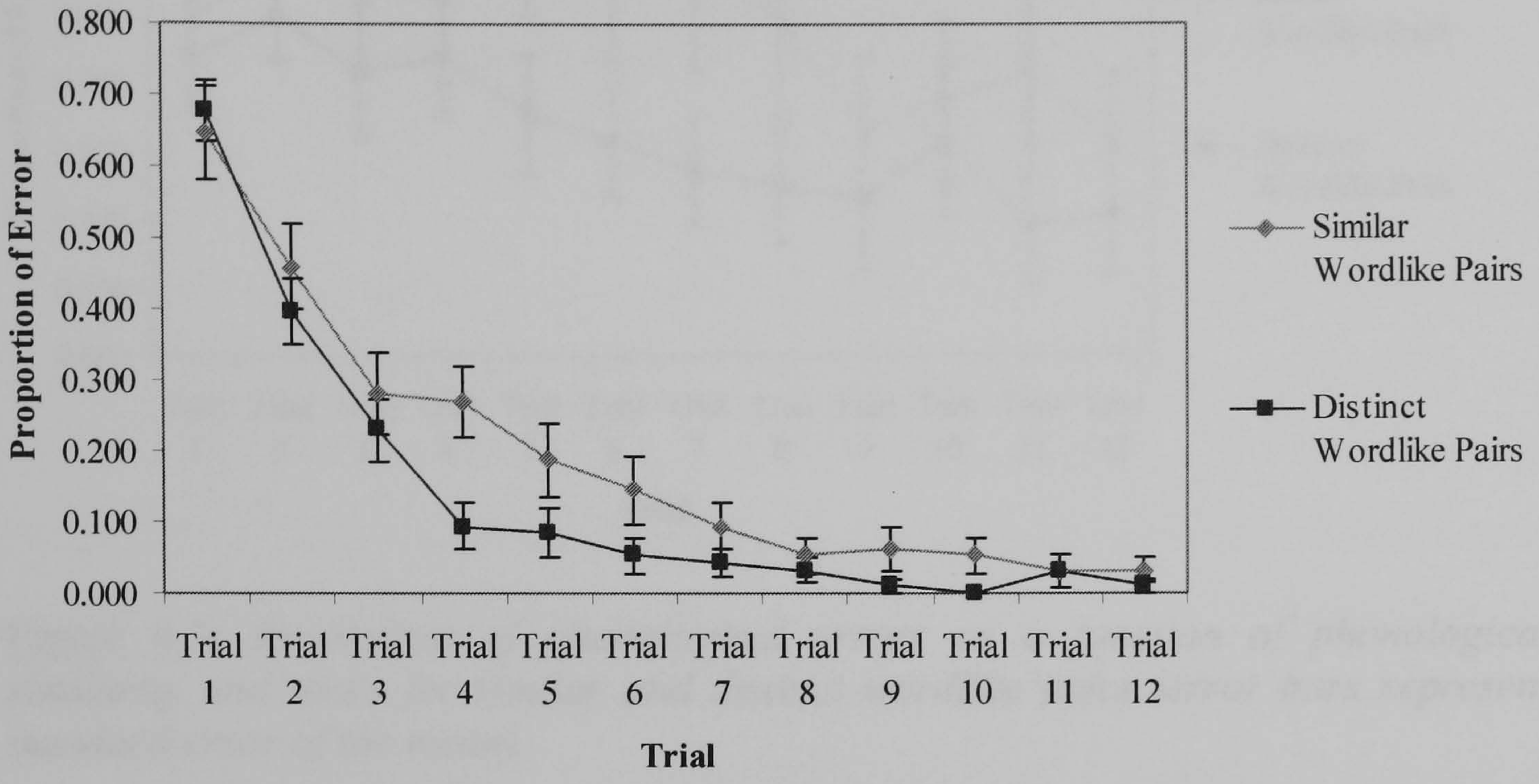


Figure 4.6: *Proportion of omissions as a function of phonological similarity and trials for similar and distinct wordlike pairs (error bars represent standard error of the mean)*

For phonological errors, the ANOVA (see Figure 4.7) reported significant main effects of phonological similarity (*F*(1,15) = 7.49, *MSe* = 0.083, *p*=.015, *r*=.58) and trials (*F*(2.800,41.999) = 3.07, *MSe* = 0.088, *p*=.041, η_p^2=.17), demonstrating a higher proportion of these errors for similar pairs, and a reduction in these errors over trials. The phonological similarity x trials interaction failed to attain significance (*F*(4.131,61.969) = 0.65, *MSe* = 0.068, *p*=.63).

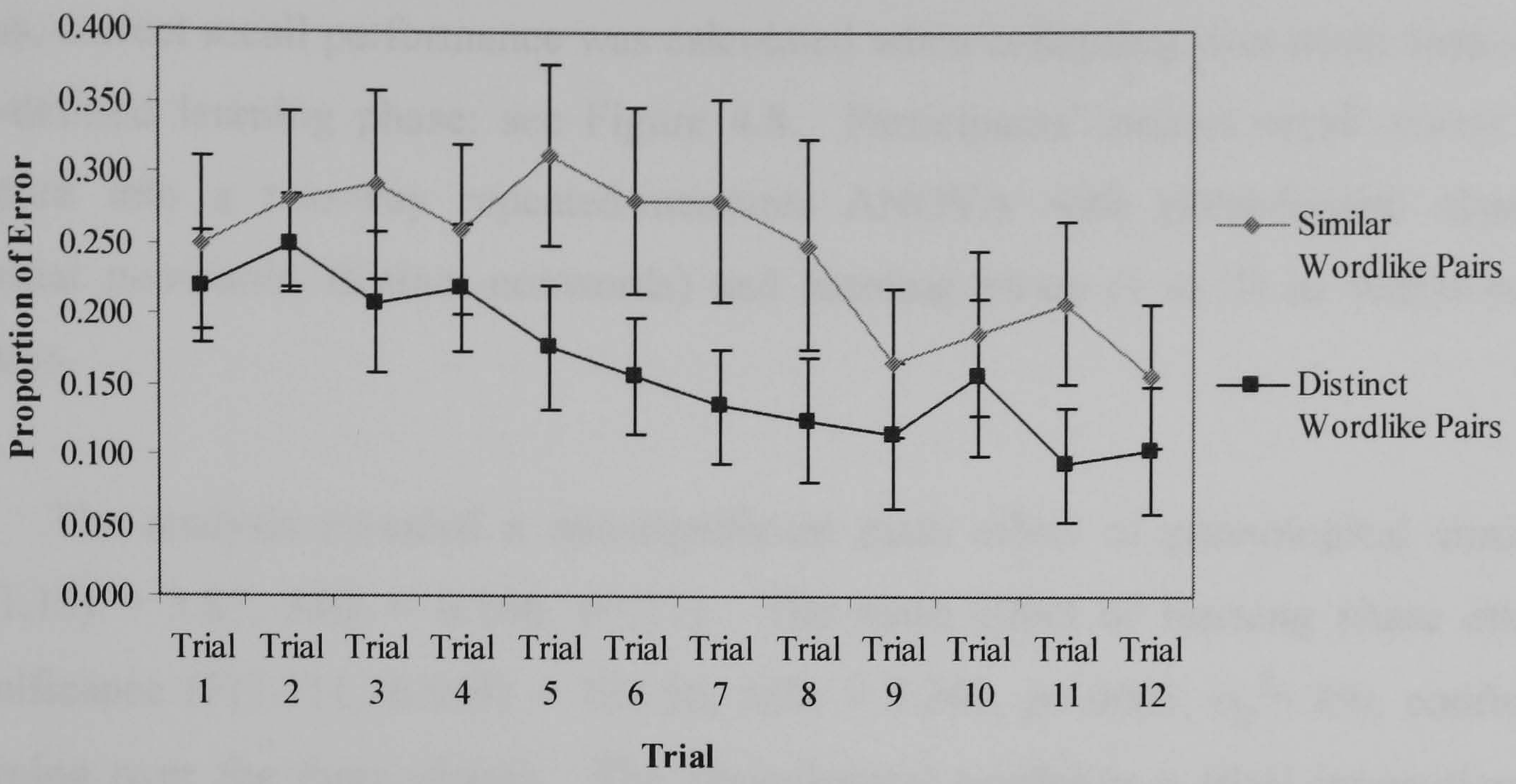


Figure 4.7: *Proportion of phonological errors as a function of phonological similarity and trials for similar and distinct wordlike pairs (error bars represent standard error of the mean)*

Word-Unwordlike Pairs: The pattern of errors revealed a higher proportion of item errors compared to omissions for both similar (means of 35% and 13%, respectively) and distinct pairs (means of 27% and 13%, respectively). The frequency of both of these error types decreased over trials. All item errors were non-lexical in nature and as such represented phonological errors. No semantically related errors or repetitions of a cue word were made. Few association errors were made (means of <4% for both similar and distinct pairs), with these occurring across trials. See Appendix 3(h).

Recall that the analysis of unwordlike pairs based on correct recall performance over all 12 trials generated a significant phonological similarity x trials interaction. This suggested that phonological similarity slowed down the learning of similar pairs at Trials 6 to 9 only, with comparable rates of learning observed for similar and distinct pairs at Trials 1 to 5, and Trials 10 to 12. It was therefore considered interesting to investigate the pattern of errors for unwordlike pairs based on this pattern of learning.

However, before conducting these error analyses, it was necessary to confirm that the phonological similarity x trials interaction observed for correct recall performance would remain when collapsing performance into these three 'phases' of learning[26]. Thus, correct recall performance was calculated when collapsing over trials within each pre-defined learning phase; see Figure 4.8. Participants' correct recall scores were entered into a two-way repeated-measures ANOVA with phonological similarity (similar nonwords, distinct nonwords) and learning phase (1 to 3) as within-subject factors.

The analysis revealed a non-significant main effect of phonological similarity ($F(1,15) = 3.87$, $MSe = 0.166$, $p=.17$). The main effect of learning phase attained significance ($F(1.111,16.669) = 123.36$, $MSe = 1.269$, $p<.0001$, $\eta_p^2=.89$), confirming learning over the three phases. The phonological similarity x trials interaction also reached significance ($F(2,30) = 11.09$, $MSe = 0.384$, $p<.0001$, $\eta_p^2=.43$), demonstrating a different rate of learning the similar and distinct unwordlike pairs over trials. An analysis of simple main effects yielded significantly better learning of the distinct unwordlike pairs at learning phase 2 ($F(1,15) = 8.63$, $MSe = 1.83$, $p=.010$, $r=.60$); performance for similar and distinct unwordlike pairs was equivalent at phases 1 and 3 (both *ps*>.05).

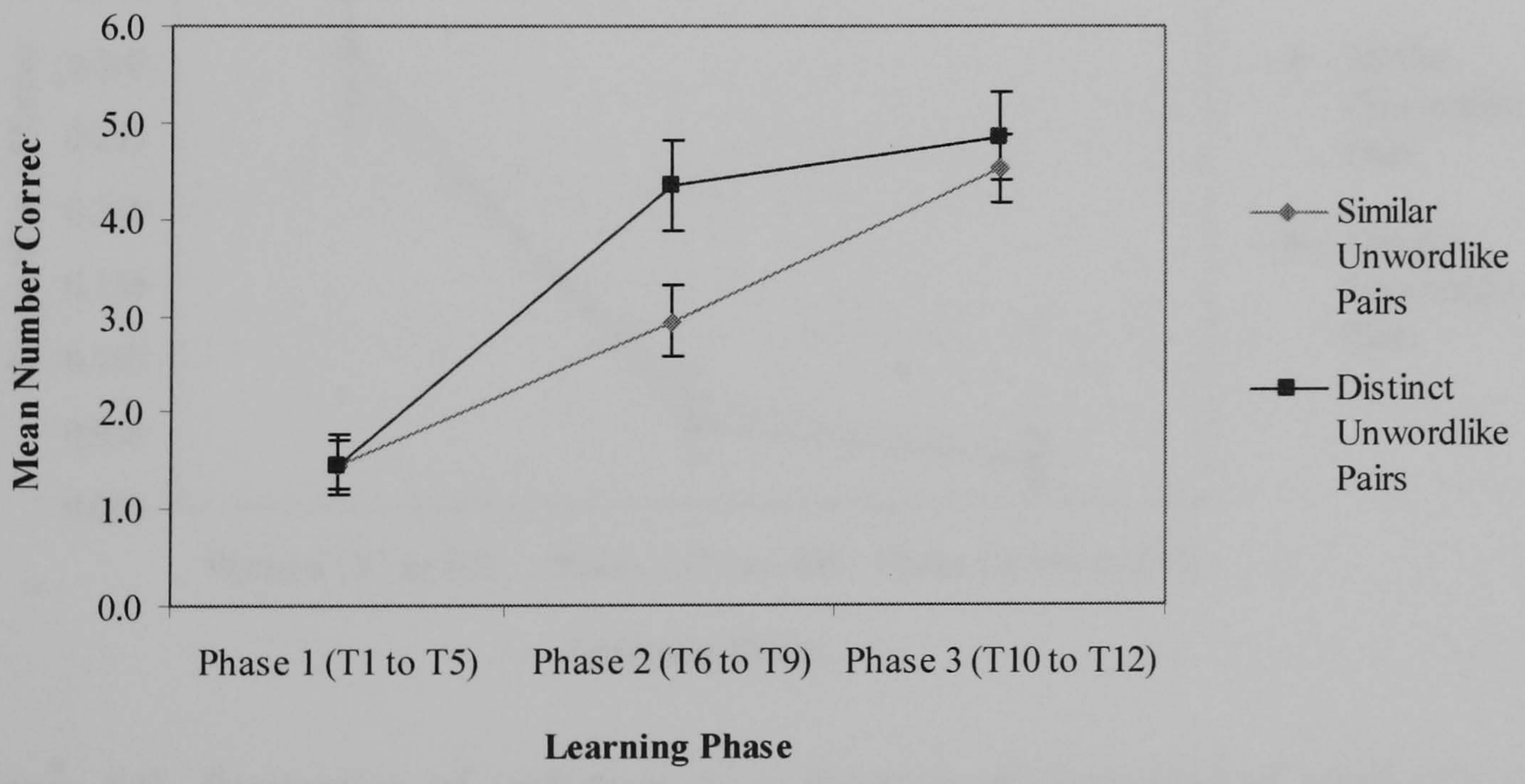


Figure 4.8: *Correct recall performance for similar and distinct unwordlike pairs when collapsing trials into three learning phases (error bars represent standard error of the mean)*

---

[26] Learning phase 1 represents performance at Trials 1 to 5; phase 2 represents performance at Trial 6 to 9; and phase 3 represents performance at Trials 10 to 12.

This revised analysis of correct recall performance generated a parallel pattern of results as reported when analysing learning over all 12 trials, confirming that phonological similarity slows down the learning of similar pairs at Trials 6 to 9 only. On the basis of these findings, each error type was subsequently collapsed into these three learning phases for the purposes of statistical analyses. Analyses were conducted on omissions and item errors (phonological errors) only. Participants' error scores, expressed as proportions, were entered into a two-way repeated-measures ANOVA incorporating the factors of phonological similarity (similar, distinct) and learning phase (1 to 3).

For omissions, the ANOVA (see Figure 4.9) revealed a main effect of learning phase ($F(1.272,19.073) = 108.99$, $MSe = 0.010$, $p<.0001$, $\eta_p^2=.88$), confirming a reduction in omissions over learning phases. The main effect of phonological similarity failed to reach significance ($F(1,15) = 0.01$, $MSe = 0.018$, $p=.93$) as did the phonological similarity x trials interaction ($F(2,30) = 1.47$, $MSe = 0.003$, $p=.25$)

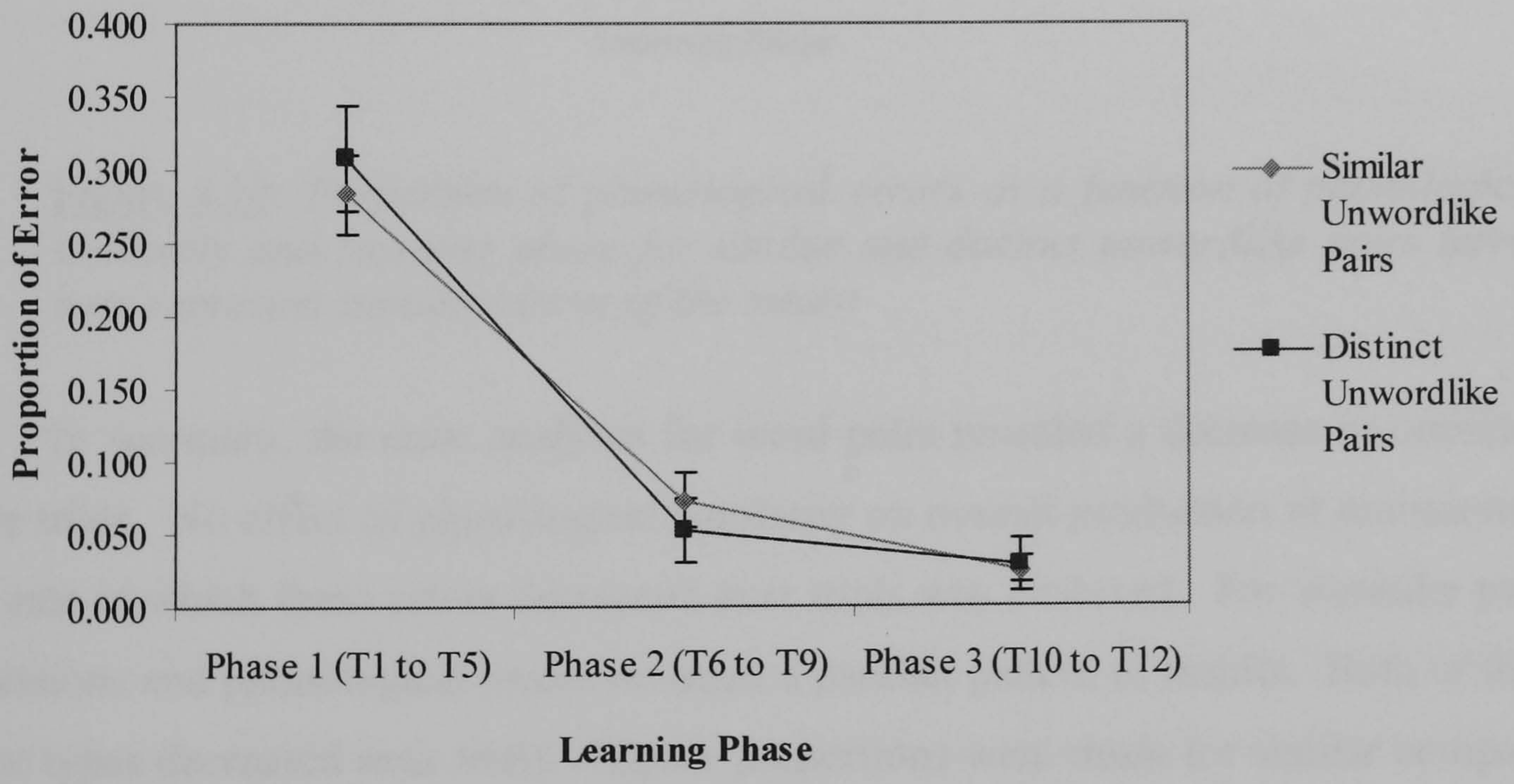


Figure 4.9: *Proportion of omissions as a function of phonological similarity and learning phase for similar and distinct unwordlike pairs (error bars represent standard error of the mean)*

For phonological errors, the ANOVA reported a non-significant main effect of phonological similarity ($F(1,15) = 2.86$, $MSe = 0.061$, $p=.11$). The main effect of learning phase attained significance ($F(1.161,17.410) = 24.89$, $MSe = 0.036$, $p<.0001$,

η_p^2=.62), confirming a reduction in these errors over learning phases. The phonological similarity x trials interaction also attained significance (*F*(2,30) = 6.74, *MSe* = 0.009, p=.004, η_p^2=.31), suggesting different learning rates over trials for similar and distinct pairs (see Figure 4.10). A simple main effects analysis yielded a significantly higher proportion of phonological errors for similar pairs at learning phase 2 (*F*(1,15) = 9.14, *MSe* = 0.03, *p*=.009, *r*=.62), with equivalent proportions at learning phases 1 and 3 (both *ps*>.05).

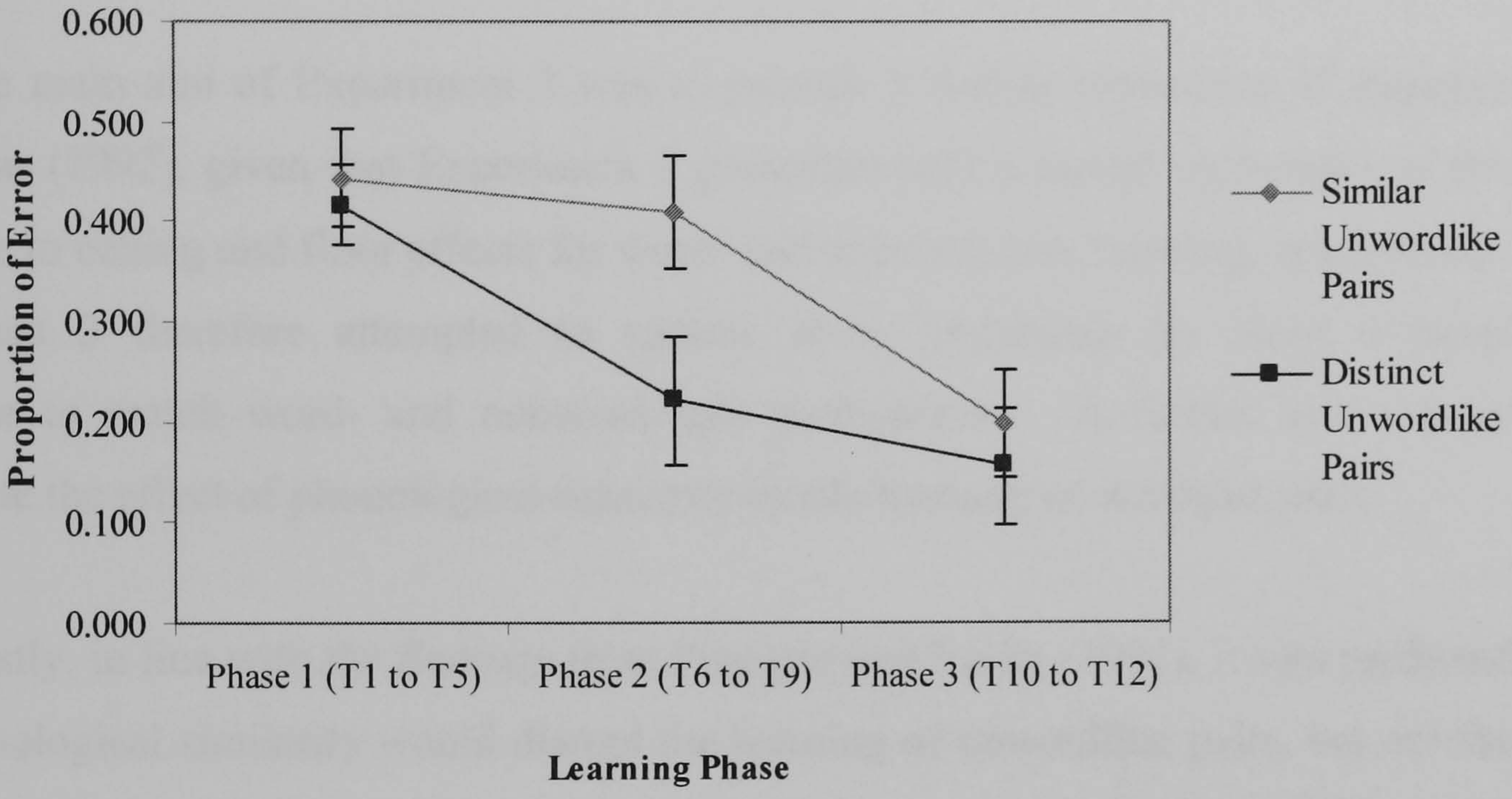


Figure 4.10: *Proportion of phonological errors as a function of phonological similarity and learning phase for similar and distinct unwordlike pairs (error bars represent standard error of the mean)*

In summary, the error analyses for word pairs revealed a decrease in omissions over trials. No effect of phonological similarity on overall production of omissions or the rate at which these errors decreased over trials was observed. For wordlike pairs, omissions and phonological errors revealed a parallel pattern of results. Both of these error types decreased over trials. Higher proportions were made for similar compared to distinct pairs for both types of errors. However, phonological similarity did not affect the rate at which these error types decreased over trials.

For unwordlike pairs, omissions and phonological errors were collapsed into three learning phases, reflecting the pattern of learning observed for correct recall performance for these pairs. Both types of errors decreased over trials. Phonological similarity did not affect overall production of these error types. For omissions, phonological similarity also failed to affect the rate at which these decreased over trials. However, phonological similarity increased the production of phonological errors; significantly more phonological errors were produced for similar pairs during the intermediate learning phase (i.e. phase 2). This latter finding parallels the pattern of learning observed for these pairs.

4.1.3 Discussion

The main aim of Experiment 3 was to provide a further replication of Papagno and Vallar (1992), given that Experiment 2 presented only a partial replication of this study due to ceiling and floor effects for word- and nonword pair learning, respectively. Experiment 3 therefore attempted to address these limitations by using a novel procedure to match word- and nonword pair performance. A further aim was to investigate the effect of phonological similarity on the learning of wordlike pairs.

Firstly, in line with the findings from Papagno and Vallar (1992), it was predicted that phonological similarity would disrupt the learning of unwordlike pairs, but not the learning of word pairs. Secondly, it was predicted that faster learning of wordlike pairs would be observed compared to unwordlike pairs. Finally, although no firm predictions were made regarding the effect of phonological similarity on the learning of wordlike pairs, it was tentatively hypothesised that phonological similarity would have a small effect on the learning of wordlike pairs when compared to the learning of unwordlike pairs.

The novel matching procedure involved extending nonword pair learning over 12 trials, with the intention of selecting the nonword pair trial (Trial N) at which performance matched that for word pairs at Trial 1. Two further design modifications were also made. Firstly, six cue-target pairs, rather than eight cue-target pairs, were presented at each trial. Secondly, cue words were replaced with more abstract words. Using the matching procedure, performance at Trial 1 for word pairs was matched with performance at Trial 3 for wordlike pairs and performance at Trial 5 for unwordlike pairs. This was confirmed by the lack of significant differences at Trial N for word, wordlike and unwordlike pairs.

Summary and discussion of word pair results

As was found to be the case in Experiment 2, the effect of phonological similarity on the learning of similar and distinct word pairs was somewhat restricted due to near-ceiling performance from Trial 3 onwards, despite the observation that performance at Trial 1 for word pairs was lower than in Experiment 2 (38%-52% in the current experiment vs. 69%-76% in Experiment 2). Thus, the current experiment failed to eliminate ceiling effects for word pairs.

Although the analysis conducted on Trials 1 and 2 revealed equivalent levels of performance for similar and distinct word pairs at both of these trials, the significant phonological similarity x trials interaction suggested faster learning of similar word pairs over these two trials. Indeed, this finding somewhat reflects Papagno and Vallar's (1992) results; they reported a significant advantage for distinct word pairs at Trial 1 with equivalent performance for similar and distinct word pairs at Trial 2. Thus, Papagno and Vallar's (1992) results may also suggest that similar word pairs are learned significantly faster from Trial 1 to Trial 2 compared to distinct word pairs. Taken together, these findings suggest that phonological similarity does not impair the learning of word pairs during the initial stages of learning.

These word pair findings offer limited support for the prediction that phonological similarity would not affect the learning of word pairs. This provides some evidence that participants do not rely on phonological coding in PSTM in order to learn word pairs, at least during the early stages of learning. Instead, given the presence of existing lexical-semantic representations of these words in LTM, participants are presumably able to utilise non-phonological codes, such as semantic codes, in order to learn familiar material. Moreover, the ceiling effects found for the learning of these pairs suggest that participants find it relatively easy to create semantic associations between the words within a pair, even when the cue words are relatively abstract. As such, the current word pairing learning results are relatively consistent with the findings from Papagno and Vallar (1992) and Experiment 2, although they are limited by the presence of ceiling effects.

Summary of nonword pair results

The effect of phonological similarity on wordlike and unwordlike pair learning was examined using two different methods of analysis. The first of these was based on the matching procedure and as such assessed learning over five trials (i.e. Trials 5 to 9 for unwordlike pairs, and Trials 3 to 7 for wordlike pairs). The second analysis took into account all 12 trials in order to examine the effect of phonological similarity over an extended period of learning. It is important to consider the patterns of nonword pair learning generated from the analysis based on the matching procedure in combination with the analysis taking into account all 12 trials in order to fully evaluate the effect of phonological similarity on nonword pair learning.

Consider first the pattern of unwordlike pair learning. The results based on learning over Trials 5 to 9 revealed that although phonological similarity impaired overall performance, the rate at which similar and distinct pairs were learned over trials was not affected by phonological similarity, as evidenced by both a non-significant phonological similarity x trials interaction and a non-significant three-way interaction. However, there was a trend in the data to suggest that phonological similarity did disrupt the learning of unwordlike pairs, with poorer performance for similar pairs at Trials 6 to 9 but equivalent performance at Trial 5. In line with this latter finding, the results based on the analysis over all 12 trials showed that phonological similarity impaired the learning of unwordlike pairs, with slower learning of similar pairs, as evidenced by the significant phonological similarity x trials interaction. However, the effect of phonological similarity in this analysis was confined to Trials 6 to 9. No differences in learning similar and distinct unwordlike pairs were observed at Trials 1 to 5 or at Trials 10 to 12. No evidence was found for a significant PSE on overall performance. Taken together, these results suggest that phonological similarity may disrupt the learning of unwordlike pairs, although the evidence is somewhat weaker when learning is based on five trials when adopting the matching procedure. However, differences were found between the two analyses regarding the effect of phonological similarity on overall performance.

Turning to the wordlike results, the analysis based on learning over Trials 3 to 7 demonstrated that phonological similarity failed to impair the learning of similar and distinct pairs, as evidenced by a non-significant phonological similarity x trials interaction and a non-significant three-way interaction. However, a PSE was observed on overall performance. The analysis based on learning over 12 trials also failed to show an effect of phonological similarity on learning similar and distinct wordlike pairs. However, there was a trend in the data to suggest that the learning of wordlike pairs was selectively disrupted by phonological similarity, with slower learning of the similar wordlike pairs. Furthermore, the negative effect of phonological similarity appears to be restricted to Trials 3 to 9, with comparable learning rates for similar and distinct pairs at Trials 1 to 2 and at Trials 10 to 12. A significant PSE was also found on overall performance. As such, it appears that there is limited evidence to suggest that phonological similarity impairs the learning of wordlike pairs. However, phonological similarity does appear to impair overall performance.

Finally, when comparing the learning of wordlike and unwordlike pairs, the results based on learning over five trials when using the matching procedure showed comparable rates of learning these two types of nonword pairs over trials. In contrast, the results based on the analysis over all 12 trials showed faster learning for wordlike compared to unwordlike pairs at Trials 2 to 6. However, comparable levels of overall performance were observed for wordlike and unwordlike pairs, although there was a trend for better performance for wordlike pairs.

The results from the analysis based on five trials using the matching procedure and the analysis based on all 12 trials do not appear to generate a consistent pattern of results for either wordlike or unwordlike pair learning. Although the matching procedure was intended to permit comparisons across the learning of word, unwordlike and wordlike pairs, it is proposed that the results based on this matching procedure are particularly limited. For example, in order to select a nonword pair trial which matched word pair performance at Trial 1, different trials were selected for the wordlike (Trial 3) and unwordlike (Trial 5) pairs. As such, nonword pair performance assessed over five trials reflects different points in learning the two types of nonword pair. As a result, such an analysis may not accurately capture any genuine effects of phonological similarity on learning wordlike and unwordlike pairs. Indeed, the finding that the negative impact of phonological similarity on nonword pair learning was restricted to particular trials based on the analysis over 12 trials further highlights how the matching procedure may mask 'true' patterns of learning wordlike and unwordlike pairs.

It is therefore argued that examining nonword pair learning over 12 trials is more informative as it provides a more accurate account of the effect of phonological similarity on learning. To this end, the remaining predictions regarding nonword pair learning will be discussed on the basis of these results.

Discussion of nonword pair results

The results for the learning of unwordlike pairs support the prediction that phonological similarity impairs the learning of unwordlike pairs, as slower learning was observed for the similar pairs. This suggests that participants utilise PSTM in order to learn unfamiliar material due to the absence of existing lexical-semantic representations of nonwords in LTM. As a result, participants are forced into relying on temporary representations of these nonwords based on their phonological structure in order to construct stable representations in LTM. These findings are in line with the results of Papagno and Vallar (1992) and to some extent the results of Experiment 2.

Of particular interest, however, is the finding that the detrimental effect of phonological similarity on learning was restricted to Trials 6 to 9; comparable rates of learning were observed at Trials 1 to 5 and Trials 10 to 12. This pattern suggests the presence of three 'phases' of learning. Indeed, this idea is further supported by the analysis conducted on correct recall performance when collapsing trials into these three phases. Phonological similarity was shown to impair learning during the intermediate phase of learning, corresponding to Trials 6 to 9; equivalent rates of learning were observed for the initial and latter phases of learning.

The lack of an effect of phonological similarity at Trials 1 to 5 cannot be entirely attributed to floor effects. Although performance at Trial 1 was at floor (0.4 and 0.1 pairs for similar and distinct pairs, respectively), learning was observed, such that performance at Trial 5 had reached 2.3 for similar pairs and 2.8 for distinct pairs. One possible suggestion regarding this pattern of learning is that participants are not utilising phonological coding during this initial phase of learning. Alternatively, it may simply be that there are too few nonword pairs in PSTM for the effect of phonological similarity to be detected. However, the finding that phonological similarity did disrupt learning at Trials 6 to 9 demonstrates that participants do utilise phonological coding to learn these pairs during an intermediate phase of learning. Finally, the finding that the negative effect of phonological similarity observed at Trials 6 to 9 disappears at later trials (Trials 10 to 12) cannot be attributed to a ceiling effect. Participants had learned only 4.3 similar and 4.7 distinct pairs (out of a maximum of 6 pairs) by Trial 12. This may imply that participants are not relying on phonological coding during this latter phase of learning. Considering this pattern of learning, it is suggested that PSTM does mediate the learning of unfamiliar material, but that the role of PSTM may change during the course of learning. Possible explanations for this varying effect of phonological similarity over trials will be discussed in detail in Chapter Seven.

The prediction that phonological similarity would have a small effect on the learning of wordlike pairs when compared to unwordlike pairs was not particularly supported. Phonological similarity was tentatively shown to impair the learning of wordlike pairs, although this evidence is not entirely supported statistically. Poorer learning of similar wordlike pairs was shown at Trials 3 to 9, with comparable learning rates observed at Trials 1 to 2 and Trials 10 to 12. This suggests the presence of three phases of learning and, as such, provides a similar pattern of learning over trials as was shown for the unwordlike pairs. Indeed, this may suggest that phonological similarity affected the learning of wordlike and unwordlike pairs in a similar manner. As was the case for unwordlike pairs, the lack of an effect of phonological similarity during the initial phase of learning may not reflect floor levels of performance given that 1.3 similar and 1.9 distinct pairs had been learned by Trial 2. Similarly, the lack of an effect of phonological similarity during the latter phase of learning (Trials 10 to 12) cannot be attributed to ceiling effects as performance had only reached 4.6 similar pairs and 5.3 distinct pairs by Trial 12. As mentioned previously, possible reasons for this pattern of results will be offered in Chapter Seven. Overall, this pattern of findings suggests that PSTM mediates, although not exclusively, the learning of wordlike pairs.

Finally, the prediction that faster learning over trials would be observed for wordlike than unwordlike pairs was supported; better learning of wordlike pairs was observed at Trials 2 to 6. This finding suggests that the learning of wordlike pairs may benefit from the existing structure of the language and is therefore consistent with Gathercole et al.'s (1996, cited in Gathercole & Martin, 1996) findings. Moreover, the finding that the effect of phonological similarity on learning emerged at an earlier trial for wordlike (Trial 3) compared to unwordlike (Trial 6) pairs may provide further evidence for faster learning of wordlike pairs. Interestingly, the finding that comparable levels of overall performance were observed for wordlike and unwordlike pairs may reflect the presence of a 'glass ceiling' for wordlike pairs. Learning of these pairs appears to slow down, particularly over Trials 9 to 12, thereby allowing the learning of unwordlike pairs to 'catch up' with the learning of wordlike pairs (0.3 vs. 0.7 pairs were learned between Trials 9 and 12 for wordlike and unwordlike pairs, respectively).

Additional analyses

An analysis of between-trial learning and forgetting rates using a Markov model could not be conducted on the present data due to the elimination of a large proportion of participants' data following the exclusion criterion. This highlights the limitations of applying this model to data with ceiling and/or floor effects, as was discussed in Chapter Three.

The analyses of errors revealed patterns which generally support the correct recall performance results. For word pairs, the predominant type of error was omissions, although these were made only infrequently. Phonological similarity did not affect the frequency with which omissions were produced, suggesting that participants were equally likely to omit a response for both similar and distinct word pairs. For wordlike pairs, phonological similarity increased the overall production of both omissions and phonological errors. This finding suggests that although participants are more likely to omit a response for similar wordlike pairs, when a response is provided it is more likely to be phonologically incorrect. However, phonological similarity did not affect the rate at which omissions and phonological errors were made over trials. For unwordlike nonwords, phonological similarity did not affect the overall production of omissions or phonological errors. However, phonological similarity did affect the rate at which phonological errors were made over trials. A higher proportion of phonological errors were made for similar unwordlike nonwords during the intermediate phase of learning (Trials 6 to 9), with equivalent proportions made for similar and distinct nonwords during the initial and latter phases of learning (Trials 1 to 5 and Trials 10 to 12). This finding suggests that similar unwordlike nonwords are particularly fragile during the intermediate phase of learning. Finally, association errors for both types of nonword pair were rare suggesting that when nonwords are recalled correctly they are associated with the correct cue word.

Taken together, these findings suggest that similar nonwords are particularly fragile and are more likely to be susceptible to errors during learning. This may imply that the syllables and/or phonemes within similar nonwords are bound together rather loosely and as a result are likely to fall apart during the course of learning. In contrast, distinct nonwords appear to be acquired more easily as a stable nonword. This conclusion is consistent with the error analyses reported in Experiment 2.

4.2 Chapter Summary

Experiment 3 generated a complex pattern of results. Phonological similarity failed to disrupt the learning of word pairs at Trials 1 and 2, suggesting that participants rely on non-phonological codes, such as semantic codes, to learn familiar material. However, similar word pairs appear to be learned at a faster rate than distinct word pairs over Trials 1 and 2. This finding is somewhat consistent with Papagno and Vallar (1992). A particularly interesting pattern of nonword pair learning emerged when providing participants with the opportunity for extended learning. Firstly, it was shown that wordlike pairs are learned faster than unwordlike pairs, suggesting that the learning of wordlike pairs may benefit from the existing structure of language. This supports previous findings (e.g. Gathercole et al., 1996, cited in Gathercole & Martin, 1996). Secondly, phonological similarity was shown to impair the learning of both wordlike and unwordlike pairs, although the evidence was somewhat weaker for the learning of wordlike pairs. This suggests that PSTM is used when learning both types of nonword pairs. That phonological similarity affected the learning of unwordlike pairs replicates the results of Papagno and Vallar (1992) and to some extent Experiment 2. Of particular interest though, was the finding that the detrimental effect of phonological similarity is restricted to an intermediate phase of learning for both wordlike and unwordlike pairs. This suggests that the role of PSTM may change during the course of learning nonword pairs. These findings can thus be seen to extend the findings of both Experiment 2 and Papagno and Vallar (1992). In line with Experiment 2, more phonological errors were produced for similar nonwords, confirming that these are more unstable and susceptible to forgetting compared with distinct nonwords. Finally, although the novel matching procedure did match nonword pair performance with word pair performance at Trial 1, it is concluded that the procedure has limited value when attempting to make comparisons between the learning of word, unwordlike and wordlike pairs given that it reflects different points in learning for the three types of pairs.

Experiment 3, and to a lesser extent Experiment 2, provided some evidence in support of the PSTM hypothesis. However, Experiment 3 reported a number of additional findings which may provide a challenge for this hypothesis. These findings will be discussed in Chapter Seven. The remainder of the experiments reported in this thesis therefore aimed to provide further tests of the PSTM hypothesis, by examining the effect of phonological similarity when adopting an alternative learning paradigm, that of the Hebb repetition task. These experiments also seek to determine whether the Hebb repetition paradigm can be viewed as an analogue to the long-term learning of novel phonological word-forms.

Chapter Five: Effects of Phonological Similarity on Long-Term Sequence Learning using the Hebb Repetition Paradigm

5.1 Introduction

Chapters Three and Four addressed the first aim of this thesis by providing further tests of the hypothesis that PSTM plays a critical role in the long-term learning of novel phonological word-forms. Experiments 2 and 3 investigated this hypothesis by conducting a replication of Papagno and Vallar (1992). The results offered some support for the hypothesis: phonological similarity was shown to selectively disrupt the learning of word-nonword pairs in Experiment 3, and to a lesser extent in Experiment 2.

The second aim of this thesis is to extend these findings to an alternative learning paradigm, that of the Hebb repetition task. The remaining experiments seek to determine whether the Hebb repetition paradigm can be viewed as an analogue to the long-term learning of novel phonological word-forms.

Chapter Two reviewed the literature pertaining to the Hebb Effect, that is, the cumulative learning of a repeated sequence over trials. The effect is robust and is found using different presentation and recall modalities and a variety of materials, such as digits, letters, words and nameable pictures (e.g. Cohen & Johansson, 1967a, 1967b; Cumming et al., 2006; Cunningham et al., 1984; Gagnon et al., 2004; Hitch et al., in press; McKelvie, 1987; Melton, 1963; Page et al., 2006; Schwartz & Bryden, 1971). However, the majority of this research has focused on the long-term learning of familiar sequences, with only a few studies examining the learning of sequences of unfamiliar material (e.g. Gagnon et al., 2004; Turcotte et al., 2005).

If the Hebb repetition paradigm is to be considered an analogue to new word-form learning, then the long-term learning of repeated sequences of familiar and unfamiliar material should reveal a corresponding pattern of results to those found in Experiments 2 and 3 and Papagno and Vallar (1992). Thus, the learning of a repeated sequence of novel phonological word-forms should be sensitive to the effects of phonological similarity. Given that novel phonological word-forms do not have existing lexical-semantic representations in LTM, participants are forced to rely on temporary representations of the phonological structure of these word-forms in order to create permanent representations in LTM. As a result, phonological similarity should disrupt the learning of a sequence of phonologically similar novel word-forms. In contrast, phonological similarity should have no effect on the long-term learning of a repeated sequence of familiar words on the basis that these words have existing lexical and semantic representations in LTM. As a result, equivalent rates of learning sequences of phonologically similar and distinct words should be observed. If parallel patterns of results are found across these two learning paradigms, this would provide converging evidence in support of the PSTM hypothesis.

However, the effects of phonological similarity on learning a sequence of familiar items may produce an alternative pattern of results. Given that the Hebb repetition task follows the format of an ISR task, phonological similarity would be expected to impair recall performance at individual trials. This is supported by the finding that phonological similarity increases the propensity of order errors (that is, recalling an item in the wrong serial position within a sequence) in ISR tasks (e.g. Conrad, 1965; Gathercole et al., 2001; Henson et al., 1996). With this in mind, learning the order of a sequence of phonologically similar words may be more laboured due to the production of these errors. If this alternative pattern of results is found for word sequence learning, this pattern will diverge from that found for word pair learning in the paired-associate task; this would suggest that learning a sequence of familiar words in the Hebb repetition task relies on PSTM.

Predictions regarding the effects of phonological similarity on long-term sequence learning in the Hebb repetition paradigm have previously been proposed by Burgess and Hitch's (1999) network model of the phonological loop (see Chapter Two for a review of this model). This model incorporates two main components: a phonological/lexical layer for phoneme and item information, and a context/timing signal for the encoding of serial order. The phonological/lexical layer is argued to be sensitive to manipulations involving phonological similarity, word length and articulatory suppression, whereas the context/timing signal is disrupted by manipulations involving serial order, such as changes in the temporal characteristics of a sequence. The model identifies the context/timing signal as the locus of the Hebb Effect (Burgess & Hitch, 1999). As such, the model predicts that sequence learning will not be disrupted by manipulations of phonological similarity, word length and articulatory suppression, given that the context/timing signal is insensitive to these phonological variables.

However, the predictions of the Burgess and Hitch (1999) model are based on the learning of a sequence of familiar items only, such as digits, letters or words. The model does not make any clear predictions about the effects of these variables on the learning of sequences of unfamiliar items, such as nonwords. The learning of a nonword presumably requires the maintenance of an ordered representation of a novel string of phonological information within the phonological/lexical layer. Thus, in contrast to learning a sequence of familiar words, learning a sequence of nonwords may be susceptible to the phonological variables of phonological similarity, word length and articulatory suppression. If this is the case, then nonword sequence learning would be disrupted by these phonological variables.

The series of experiments presented in the remainder of this thesis serve to investigate the effects of Hebb repetition and phonological similarity on the learning of word and nonword sequences.

5.2 Experiment 4: Effects of Phonological Similarity and Hebb Repetition on Word and Nonword Sequence Learning

Experiment 4 represents the first attempt to investigate the effects of phonological similarity and Hebb repetition on the long-term learning of word and nonword sequences. The design follows that of the original Hebb experiment (Hebb, 1961) with the repeated Hebb sequence presented every third trial, with two intervening nonrepeated filler sequences. A number of criteria were adopted for the detection of a Hebb Effect with a distinction being made between strong and weak evidence for a Hebb Effect. Firstly, a clear and reliable Hebb Effect would be indicated by a main effect of list type, demonstrating better overall performance for Hebb than filler sequences, coupled with a list type x trials interaction, indicating cumulative learning over trials for the Hebb sequence over and above any general practice effect observed for the filler sequences. Secondly, weaker evidence for a Hebb Effect would be indicated by a main effect of list type but no list type x trials interaction or vice versa.

Thirdly, weak evidence for a Hebb Effect would be indicated by a main effect of trials in the absence of a main effect of list type and a list type x trials interaction. Finally, the absence of a Hebb Effect would be indicated by a simultaneous failure to obtain a main effect of list type, a list type x trials interaction and a main effect of trials.

As mentioned previously, few studies have investigated the Hebb Effect for sequences of nonwords (e.g. Gagnon et al., 2004; Turcotte et al., 2005). In these studies, participants were presented with nonword sequences at their span plus one item. This method ensures that all participants are presented with nonword sequences exceeding their individual STM capacities, and are in this sense matched on STM ability. Furthermore, this method may increase the likelihood of obtaining a Hebb Effect as participants would be expected to easily learn a sequence which is only one item longer than their STM capacity over the course of repeated presentations. Indeed, Gagnon et al. (2004) and Turcotte et al. (2005) both showed evidence of Hebb Effects for nonword sequences. However, although this method may facilitate sequence learning, there is the risk that learning may reach ceiling before the final presentation of the repeated sequence. Indeed, such a finding was shown by Turcotte et al. (2005). These authors reported an average nonword sequence span of 4 for a group of young adults (mean age of 22 years); based on this span an average sequence length of 5 nonwords would have been adopted for the Hebb repetition task. The authors found nonword sequence learning to have reached 75% correct by the fourth presentation of the repeated sequence, with performance increasing to 85% correct by the seventh presentation. The Hebb Effect in this study was therefore shown by only a 10% increase over presentations[27]. To this end, in order to avoid ceiling performance in the current experiment, nonword sequence length was set at six items for all participants.

A number of predictions were made. Firstly, it was predicted that Hebb Effects would emerge for word and nonword sequences. Secondly, phonological similarity was predicted to impair ISR performance on individual trials for both word and nonword sequences. Finally, based on Papagno and Vallar's (1992) paired-associate results, it was hypothesised that equivalent learning rates would be observed for similar and

---

[27] Sequence learning was compared over two time points only: the first was an average of the second to fourth presentations of the repeated sequence; the second was an average of the fifth to seventh presentations of the repeated sequence.

distinct word sequences. In contrast, it was predicted that phonological similarity would selectively impair the learning of repeated nonword sequences; that is, significantly faster learning was expected for distinct compared to similar nonword sequences. The predictions regarding word sequence learning also conform to those advanced by the Burgess and Hitch (1999) model.

5.2.1 Method

5.2.1.1 Design

Participants undertook an ISR task in which sequences of phonologically similar and phonologically distinct words and nonwords were auditorily presented; participants were then required to immediately recall the items in their correct serial order. One sequence was repeated every third trial (the Hebb list) interspersed with two non-repeated trials (filler lists). The Hebb list and each of the two filler lists each comprised a different set of items. Items in the Hebb list were presented in the same serial position on each trial and items in every filler list were presented in a different serial position on each trial.

The experiment followed a 2 x 2 x 8 x 2 x 2 mixed-design incorporating three within-subjects factors: phonological similarity (similar items, distinct items), list type (Hebb, filler) and trials (1 to 8); and two between-subjects factors: lexicality (words, nonwords) and materials (materials A, materials B). The lexicality and phonological similarity factors combined to produce four experimental conditions: similar word sequences; distinct word sequences; similar nonword sequences; and distinct nonword sequences.

Participants were randomly assigned to one of the lexicality conditions. Within each lexicality condition, the factor of phonological similarity was counterbalanced across participants so that half received similar sequences followed by distinct sequences, and half the reverse. As in the previous experiments, two sets of materials (A and B) were constructed for each of the four experimental conditions. Participants were randomly divided so that half received materials A, and half materials B. For each set of materials within each experimental condition, three different sets of items were devised: one for the repeated Hebb list (H) and one for each of the two non-repeated filler lists (Filler 1, F_1; Filler 2, F_2). Lists were randomly designated as F_1, F_2 or H. The presentation order of the list type factor was F_1, followed by F_2, followed by H. For each set of materials within each of the experimental conditions, the order of items within a list remained the same across participants.

5.2.1.2 Participants

Thirty-two undergraduate students attending the University of York volunteered to participate for either course credit or payment. There were 25 females and 7 males aged between 18 years and 22 years (mean age of 19.8 years). All participants spoke English as their native language and reported no known hearing or language impairments.

5.2.1.3 Apparatus

The same apparatus were used as in Experiments 1 to 3, with the exception that participants heard the materials via a pair of Beyerdynamic DT 770 headphones.

5.2.1.4 Materials

The materials were all monosyllabic 3-phoneme CVC words and nonwords. Each experimental set contained either seven words or six nonwords; the smaller size of the nonword sets was an attempt to equate task difficulty, as in Experiment 1.

Word Sets

Twelve sets of seven words were constructed: six phonologically similar sets (three sets for each of materials A and B) and six phonologically distinct sets (three sets for each of materials A and B). Each of the 12 sets were matched on written frequency per million (Kucera & Francis, 1967; range across sets: 7.1 to 11.4) using the MRC Psycholinguistic Database (Coltheart, 1981) and neighbourhood size[28] (range across

---

[28] Neighbourhood size was based on the number of monosyllabic 3-phoneme words in the CELEX database (Baayen et al., 1993) differing from the target word by the substitution of a single phoneme (i.e.

sets: 26.0 to 31.6) using the CELEX psycholinguistic database (Baayen, Piepenbrock & van Rijn, 1993). Two separate one-way between-subjects ANOVAs revealed no statistically significant differences between the 12 word sets for written frequency ($F(11,72) = 0.13$, $MSe = 102.611$, p=1.00) or neighbourhood size ($F(11,72) = 0.81$, $MSe = 36.218$, p=.63). See Appendix 4(a) for the word sets.

Phonologically Similar Word Sets: These were constructed so that the words had similar sounding phonological representations (but did not all rhyme). Within a set, each word had the same vowel and used two different initial consonants (C_1) and five different final consonants (C_2)[29] (e.g. *rat, ram, rap, rag, cab, cat, cam*). Three sets used the vowel */a/* (materials A) and three used the vowel */u/* (materials B).

Phonologically Distinct Word Sets: These were constructed so that the words had distinct sounding phonological representations. Each set used four different vowels (excluding the vowels */a/* for materials A and */u/* for materials B). Each word within a set used a different C_1 and a different C_2 (e.g. *vet, tip, jug, pod, bin, gum, lob*)[30].

Nonword Sets

Twelve sets of six nonwords were created: six phonologically similar sets (three sets for each of materials A and B) and six phonologically distinct sets (three sets for each of materials A and B). Each of the sets was constructed following the same criteria as were adopted for the construction of the phonologically similar and distinct word sets. Each nonword failed to match a lexical entry within the CELEX database (Baayen et al., 1993). Each of the 12 sets was matched on neighbourhood size using the CELEX database (mean range across sets: 10.5 to 13.7). A one-way between-subjects ANOVA revealed no statistically significant differences between the 12 nonword sets ($F(11,60) = 0.26$, $MSe = 17.742$, p=.99). See Appendix 4(b) for the nonword sets.

$C_1V\_$, $\_VC_2$ and $C_1\_C_2$). These were then summed to obtain an estimate of neighbourhood size (e.g. Thorn & Frankish, 2005).

[29] This procedure closely followed that adopted by Baddeley (1966a).

[30] One list used six different C_2.

Phonologically Similar Nonword Sets: Within a set, each nonword had the same vowel and used two different C_1 and four different C_2 (e.g. *bem, bep, bez, nef, nep, nez*). Three sets used the vowel */e/* (materials A) and three used the vowel */o/* (materials B).

Phonologically Distinct Nonword Sets: Each set used five different vowels[31]. Each nonword within a set used a different C_1 and a different C_2 (e.g. *zep, fub, nid, heg, sof, vas*).

Word Sets vs. Nonword Sets

Mean (SD) neighbourhood size values were 27.7 (5.9) and 12.1 (4.0) for word and nonword sets, respectively. A two-tailed independent t-test revealed a statistically significant difference between these two types of set ($t(145.624) = 19.56$, $p<.0001$, $r=.85$), confirming smaller neighbourhood size values for nonword sets.

Editing of Speech Signals

Each item was recorded onto a minidisc player in a soundproof room by the same speaker as used in Experiments 1 to 3. Each item was edited following the same procedure as described in Experiment 1.

5.2.1.5 Procedure

The ISR task was conducted with individual participants in a quiet testing room. Participants were informed the experiment would assess their short-term memory for sequences of words or nonwords and would take approximately 30 mins. The instructions advised participants they would hear sequences containing either seven words or six nonwords which they would be required to immediately recall aloud in exactly the same order in which they were presented. Participants were instructed to say 'blank' if they could not remember the item that appeared in a particular position in the sequence. Participants were further advised that beeps would signal the presentation of each sequence and the beginning of each recall period. Participants were informed that they would be presented with two blocks of 24 sequences. Participants were made

---

[31] The use of four vowels within the distinct word lists as opposed to the use of five vowels within the distinct nonword lists was the result of an experimental design error. This error was corrected in future experiments.

aware that their responses would be recorded onto a minidisc player in order to facilitate scoring and so they were to articulate their responses as clearly as possible.

Two practice trials, each containing seven words or six nonwords not present in the experimental trials, were given prior to presenting the experimental sequences in order to check the volume level and to allow participants to become accustomed to the procedure and the speaker's voice.

Each experimental sequence began with a 1.5 s pause, in which a short beep was presented, followed by the seven words or six nonwords. Items were presented at a rate of 1 s per item. After the presentation of the final item within a sequence, a 1.5 s pause, again including a small beep, was presented indicating the beginning of the spoken recall period. When participants indicated that they were unable to recall any further items from the sequence, the next sequence was presented. A 1 min interval occurred between the two blocks of 24 sequences.

Responses were scored both immediately using a strict serial recall criterion and later using the minidisc recordings to verify scoring and to transcribe the exact pronunciation of each item, particularly the nonwords. Maximum scores for each list were 7 for word sequences and 6 for nonword sequences.

5.2.2 Results

5.2.2.1 Correct Recall Analysis

Responses were scored following the same strict serial recall criterion as described in Experiment 1. No instances occurred in which a participant consistently made the same mispronunciation of an item. Instances in which the correct pronunciation of an item could not be clearly determined were scored as correct, provided they were recalled in the correct serial position; this occurred on two occasions[32].

---

[32] The word '*bum*' was perceived as '*bomb*', presumably due to accent differences between the participant and experimenter; the target nonword '*sem*' was difficult to determine from the incorrect nonword '*zem*'.

The data were normally distributed and no outliers were detected[33]. Inter-rater reliability analysis was conducted by an independent rater[34] on a random sample of 19% (n=6) of the participants' responses. This revealed a perfect correspondence for each of the four conditions.

For each lexicality condition, the means and standard deviations were firstly calculated for F_1, F_2, and H based on the eight participants receiving materials A and the eight participants receiving materials B. In line with Experiment 3, to avoid conducting a complex five-way ANOVA for the main statistical analysis, these data were firstly subjected to a number of separate three-way mixed-design and repeated-measures ANOVAs in order to confirm parallel patterns of results across (i) materials A and B for each of F_1, F_2 and H for each lexicality condition; and (ii) F_1 and F_2 for each lexicality condition[35]. These analyses were initially conducted over all eight trials; however, error variances for nonword sequence learning varied greatly. Thus, in an attempt to reduce error variance for nonword sequence learning, thereby improving both statistical power and sensitivity of analyses, mean scores were averaged over pairs of trials (i.e. Trials 1 and 2; Trials 3 and 4; and so on). Sequence learning was therefore assessed over pairs of successive presentations[36]. As a result, sequence learning was examined over Trials 1 to 4, where Trial 1 represents the average of Trials 1 and 2, Trial 2 represents the average of Trials 3 and 4, Trial 3 represents the average of Trials 5 and 6, and Trial 4 represents the average of Trials 7 and 8. This procedure applies to all remaining experiments presented in this thesis.

The majority of these analyses revealed parallel patterns of results, allowing the data to be pooled across materials A and B, and across F_1 and F_2. However, a cautionary note needs to accompany the outcome of all further statistical analyses given that two significant interactions were reported (see Appendix 4(c) for a summary of

---

[33] Outliers were defined as in Experiment 1 and were calculated separately for F_1, F_2 and H for each experimental condition in each of materials A and B.

[34] The independent rater was the same as in Experiments 1 and 2 and had not taken part in Experiments 4 to 7. Of the six sets of participants' responses scored, three had completed the word conditions and three the nonword conditions.

[35] It was necessary to confirm parallel patterns of performance for F_1 and F_2 so that an average of these two lists could be taken in order to compare performance on filler lists with that of Hebb lists.

[36] This procedure of averaging over trials has been previously adopted by studies assessing sequence learning via the Hebb repetition task in order to "create a more stable estimation of performance" (Turcotte et al., 2005, p. 253)

these analyses). Despite these interactions, it is important to clarify that the combined analysis did not misrepresent the results in any major way. Thus, the means and standard deviations were therefore calculated based on the 16 participants in each of the two lexicality conditions.

As in Experiments 2 and 3, the word and nonword data were analysed separately in order to avoid conducting a complex four-way analysis. Participants' scores were entered into two separate three-way repeated-measures ANOVAs each incorporating three within-subjects factors: phonological similarity (similar items, distinct items), list type (Hebb sequences, filler sequences) and trials (1 to 4).

The ANOVA conducted on the word data revealed statistically significant main effects of phonological similarity ($F(1,15) = 22.19$, $MSe = 4.142$, $p<.0001$, $r = .77$) and list type ($F(1,15) = 26.98$, $MSe = 2.255$, $p<.0001$, $r=.80$), confirming better recall performance for distinct word sequences, and for Hebb lists. The main effect of trials also attained significance ($F(3,45) = 24.40$, $MSe = 0.736$, $p<.0001$, $\eta_p^2=.62$), indicating learning over trials (see Figure 5.1).

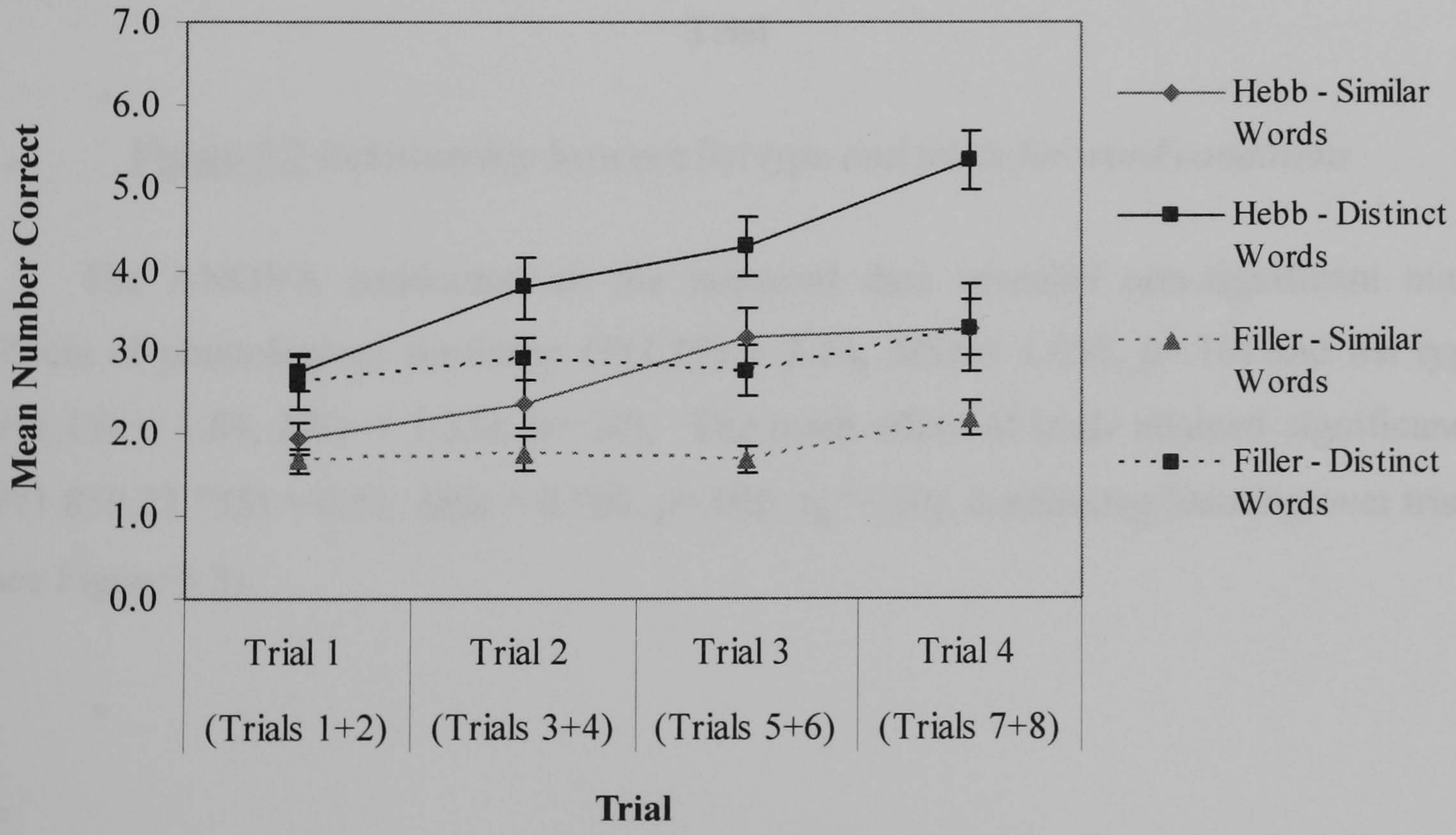


Figure 5.1: *ISR performance for word conditions as a function of phonological similarity, list type and trials (error bars represent standard error of the mean)*

The list type x trials interaction reached significance ($F(3,45) = 11.65$, $MSe = 0.593$, $p<.0001$, $\eta_p^2=.44$), indicating a different pattern of learning over trials for Hebb and filler lists (see Figure 5.2). A simple main effects analysis yielded significant learning over trials for Hebb lists ($F(3,45) = 23.21$, $MSe = 0.98$, $p<.0001$) and filler lists ($F(3,45) = 6.16$, $MSe = 0.35$, $p=.001$). Thus, this interaction demonstrates faster learning over trials for the Hebb compared to filler lists. All remaining interactions failed to reach significance (all *p*s>.05).

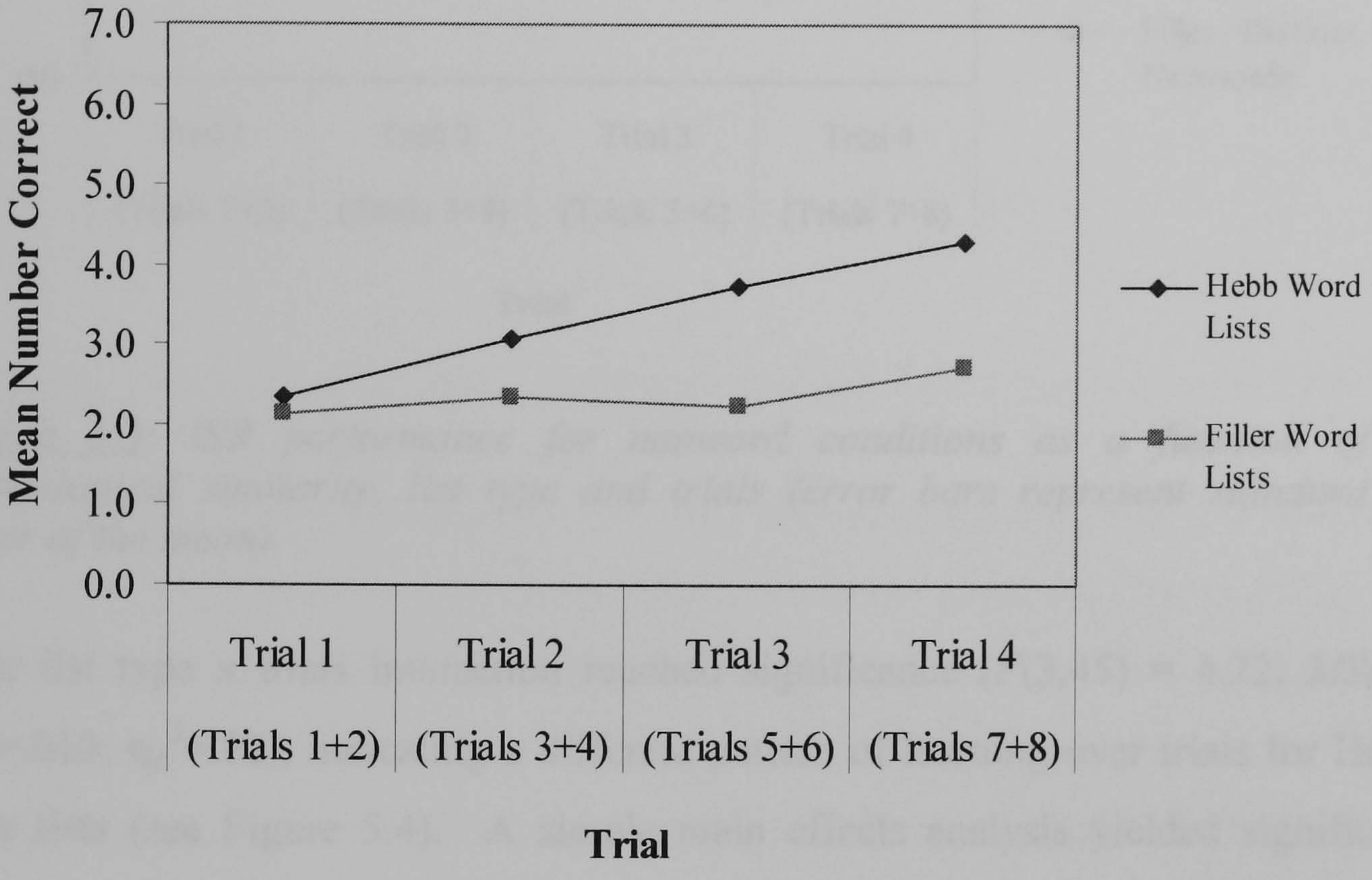


Figure 5.2 *Relationship between list type and trials for word conditions*

The ANOVA conducted on the nonword data revealed non-significant main effects of phonological similarity ($F(1,15) = 2.24$, $MSe = 1.658$, $p=.16$) and list type ($F(1,15) = 1.84$, $MSe = 1.351$, $p=.20$). The main effect of trials attained significance ($F(1.850,27.755) = 6.51$, $MSe = 0.709$, $p=.006$, $\eta_p^2=.30$), confirming learning over trials (see Figure 5.3).

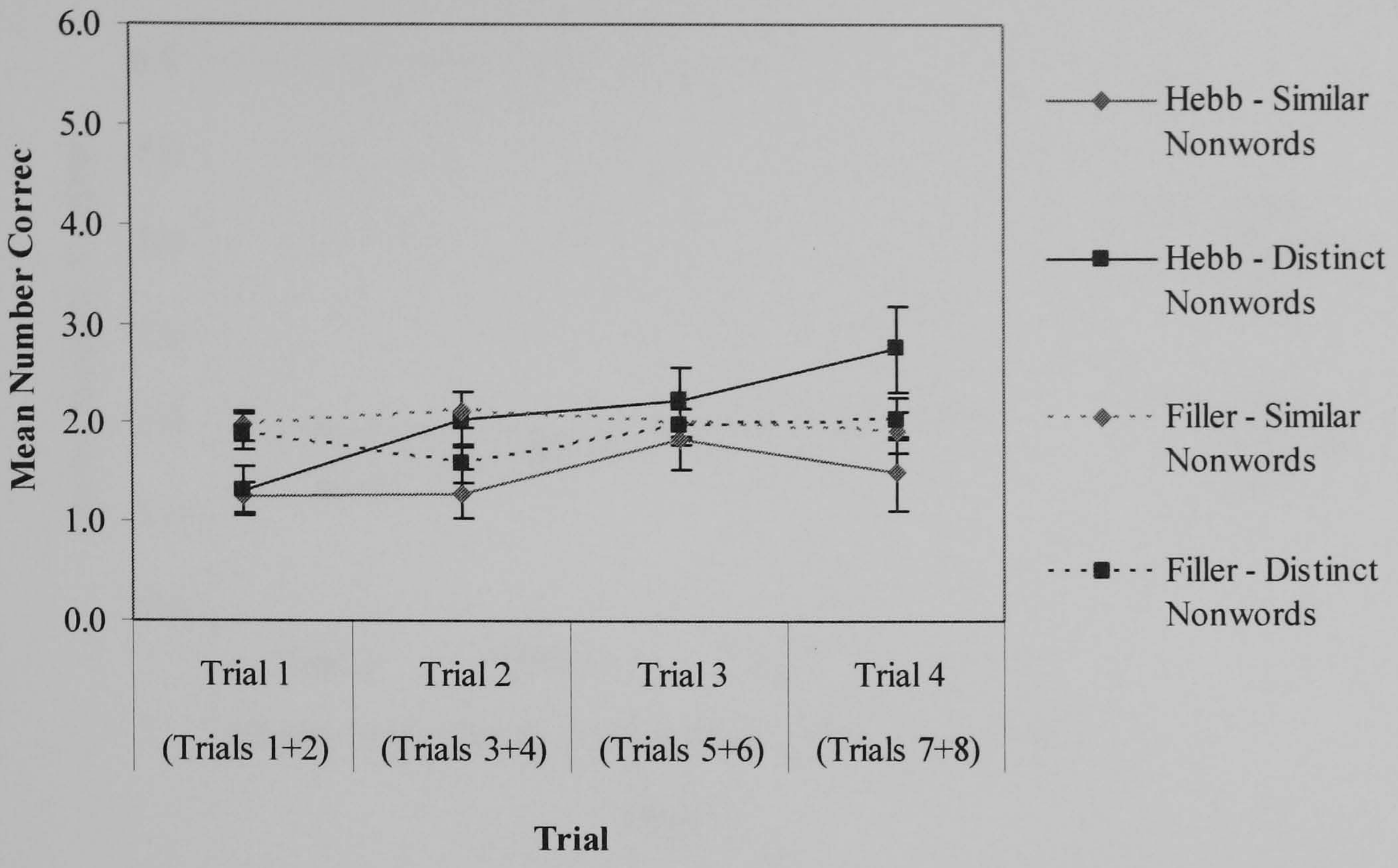


Figure 5.3: *ISR performance for nonword conditions as a function of phonological similarity, list type and trials (error bars represent standard error of the mean)*

The list type x trials interaction reached significance ($F(3,45) = 4.22$, $MSe = 0.478$, $p=.010$, $\eta_p^2=.22$), indicating a different pattern of learning over trials for Hebb and filler lists (see Figure 5.4). A simple main effects analysis yielded significant learning over trials for Hebb lists ($F(3,45) = 7.33$, $MSe = 0.65$, $p<.0001$) but not for filler lists ($F(3,45) = 0.38$, $MSe = 0.27$, $p=.77$).

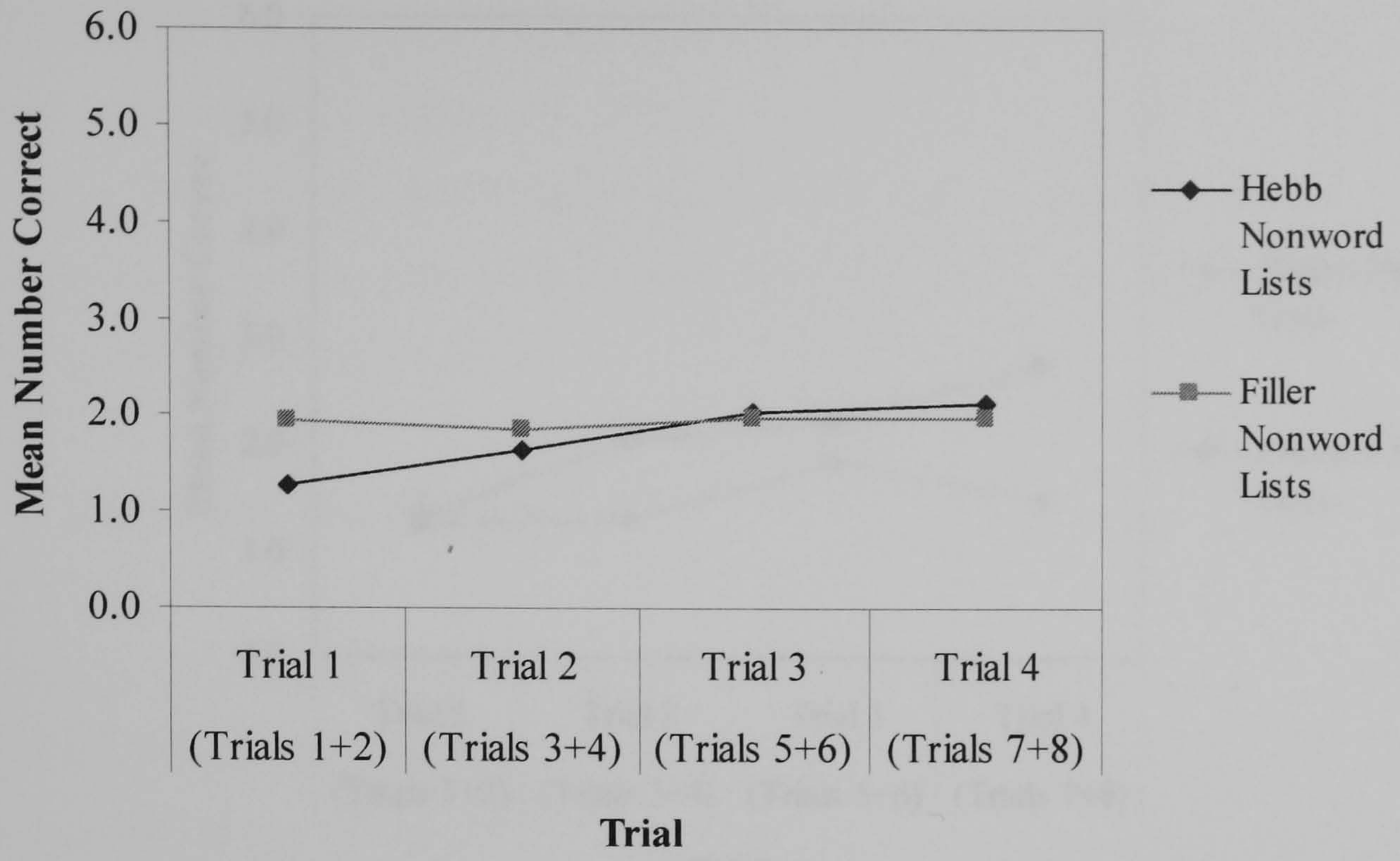
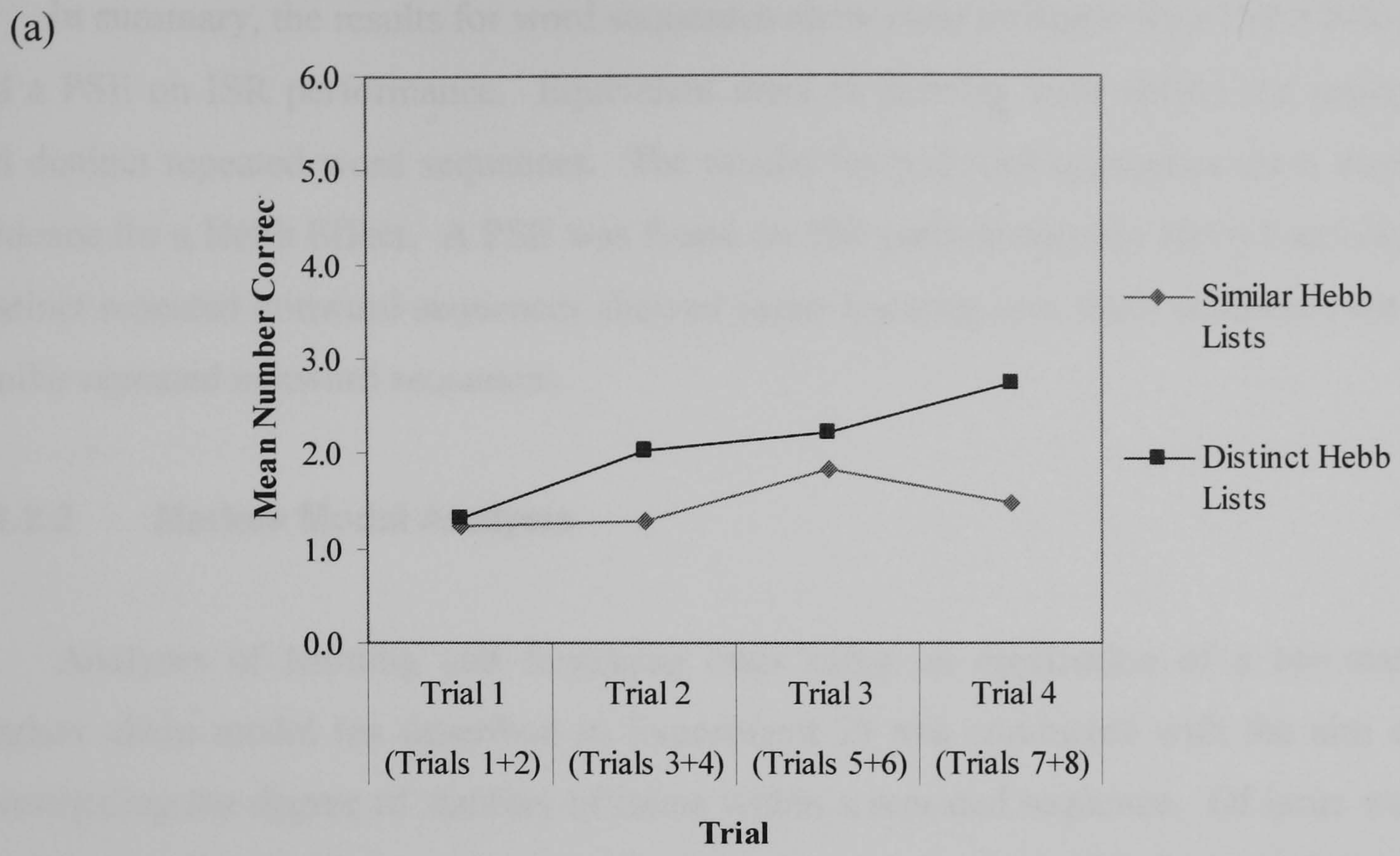


Figure 5.4 *Relationship between list type and trials for nonword conditions*

Significant two-way interactions emerged between phonological similarity and list type ($F(1,15) = 6.12$, $MSe = 1.423$, $p=.026$, $r=.54$) and phonological similarity and trials $F(3,45) = 2.90$, $MSe = 0.510$, $p=.045$, $\eta_p^2=.16$). Moreover, these were incorporated into a significant three-way interaction ($F(3,45) = 3.95$, $MSe = 0.318$, $p=.014$, $\eta_p^2=.21$). In order to refine the extent of the effect of phonological similarity on learning over trials, the data from each list type were submitted independently to 2 x 4 repeated-measures ANOVAs with phonological similarity and trials as factors (see Figure 5.5).

For the Hebb list data, significant main effects of phonological similarity ($F(1,15) = 4.63$, $MSe = 2.566$, $p=.048$, $r=.49$) and trials ($F(1.713,25.697) = 7.33$, $MSe = 1.137$, $p=.004$, $\eta_p^2=.33$) emerged, demonstrating better recall performance for distinct Hebb lists, and learning over trials. The phonological similarity x learning trials interaction attained significance ($F(3,45) = 4.46$, $MSe = 0.469$, $p=.008$, $\eta_p^2=.23$), suggesting a different pattern of learning over trials for similar and distinct nonword sequences. A simple main effects analysis yielded significant learning over trials for distinct nonword sequences ($F(3,45) = 10.20$, $MSe = 0.55$, $p<.0001$) but not for similar nonword sequences ($F(3,45) = 2.12$, $MSe = 0.56$, $p=.11$).



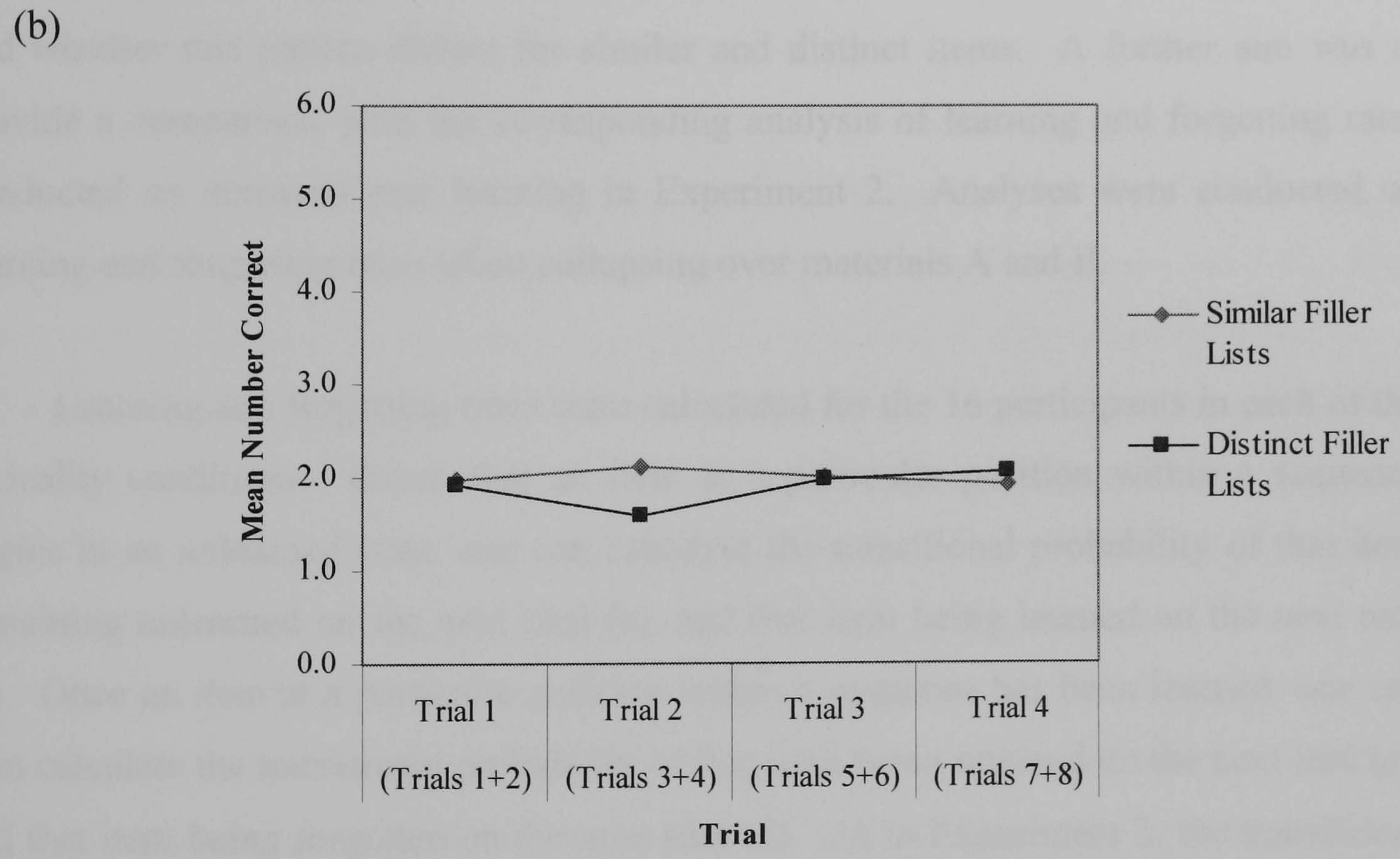


Figure 5.5: *Relationship between phonological similarity and trials for (a) Hebb lists; and (b) filler lists*

For the filler list data, the ANOVA reported non-significant main effects of phonological similarity (*F*(1,15) = 1.02, *MSe* = 0.514, *p*=.33) and trials (*F*(3,45) = 0.38, *MSe* = 0.265, *p*=.77). The phonological similarity x trials interaction failed to reach significance (F(3,45) = 1.80, MSe = 0.360, p=.16).

In summary, the results for word sequences show clear evidence for a Hebb Effect and a PSE on ISR performance. Equivalent rates of learning were shown for similar and distinct repeated word sequences. The results for nonword sequences show weak evidence for a Hebb Effect. A PSE was found on ISR performance for Hebb lists only. Distinct repeated nonword sequences showed faster learning over trials compared with similar repeated nonword sequences.

5.2.2.2 Markov Model Analysis

Analyses of learning and forgetting rates using an application of a two-state Markov chain model (as described in Experiment 2) was conducted with the aim of investigating the degree of stability of items within a repeated sequence. Of issue was whether it is the case that once an item has been learned, in terms of both the item itself (i.e. the sequence of phonemes) and its position within a sequence, it remains learned; and whether this pattern differs for similar and distinct items. A further aim was to provide a comparison with the corresponding analysis of learning and forgetting rates conducted on nonword pair learning in Experiment 2. Analyses were conducted on learning and forgetting rates when collapsing over materials A and B.

Learning and forgetting rates were calculated for the 16 participants in each of the lexicality conditions. Given that an item at a particular position within a sequence begins in an unlearned state, one can calculate the transitional probability of that item remaining unlearned on the next trial (a), and that item being learned on the next trial (b). Once an item at a particular position within a sequence has been learned, one can then calculate the transitional probability of that item being retained on the next trial (c), and that item being forgotten on the next trial (d). As in Experiment 2, the transitional probabilities (b) and (d) represent learning and forgetting rates, respectively

For each participant, the transitional probabilities (b) and (d) were calculated for each of seven transitional steps: Trial 1 to Trial 2 (T1-T2), Trial 2 to Trial 3 (T2-T3), Trial 3 to Trial 4 (T3-T4), Trial 4 to Trial 5 (T4-T5), Trial 5 to Trial 6 (T5-T6), Trial 6 to Trial 7 (T6-T7), and Trial 7 to Trial 8 (T7-T8) for the Hebb lists only[37], following the

---

[37] Learning and forgetting rates were not calculated for filler lists as the items in these lists were presented in a different serial order at each presentation.

same strict criteria as described in Experiment 2. Calculation of transitional probabilities took into account all seven words or all six nonwords at each transitional step. Instances arose in which transitional probabilities (b) and (d) could not be calculated at some (or all) transitional steps. Failure to obtain (b) reflected situations in which all items within a sequence were recalled in the correct position on one trial and remained in the correct position on the next trial. Situations in which (d) could not be calculated reflected instances in which no items within a sequence were recalled in the correct position on one trial, and no items within a sequence being recalled in the correct serial position on the next trial. Participants were excluded if transitional probabilities (b) and/or (d) could not be calculated for one (or more) transitional steps. Following these criteria, missing data resulted in the exclusion of 75% (n=12) of participants from the word conditions and 94% (n=15) of participants from the nonword conditions.

To reduce the amount of missing data, participants' transitional probabilities were averaged over the seven transitional steps. Although this method results in a loss of information about learning and forgetting rates on a trial-by-trial basis, there remains the potential to gain an insight into the nature by which phonological similarity may affect how a sequence of items is learned. Participants' data were excluded if an averaged transitional probability could not be calculated due to missing data at all seven transitional steps in any one of the two conditions. No participants were excluded following this revised criterion. Averaged transitional probabilities were therefore calculated based on all 16 participants in each lexicality condition.

A differential pattern of results was observed for the word and nonword data (see Table 5.1). Phonological similarity appears to affect both learning and forgetting rates for word sequences. A higher learning rate was observed for distinct word sequences than similar word sequences. In contrast, a higher forgetting rate was found for similar compared to distinct word sequences. For the nonword conditions, phonological similarity appears to selectively affect forgetting rates only. A higher forgetting rate was observed for similar compared to distinct nonword sequences. Comparable learning rates were found for similar and distinct nonword sequences.

Table 5.1: *Mean (and standard deviation) transitional probabilities for learning and forgetting transition types for similar and distinct word and nonword sequences*

	Condition			
	Word Sequences		Nonword Sequences	
Transition Type	Similar	Distinct	Similar	Distinct
Learning	.238 (.137)	.381 (.196)	.201 (.138)	.207 (.120)
Forgetting	.345 (.201)	.169 (.128)	.598 (.276)	.373 (.296)

In line with the analyses based on correct recall performance, the word and nonword data were analysed separately. Participants' averaged transitional probability values were entered into two-way repeated-measures ANOVAs each incorporating two within-subject factors: transition type (learning, forgetting) and phonological similarity (similar, distinct).

The ANOVA conducted on the word data reported non-significant main effects of phonological similarity (*F*(1,15) = 0.43, *MSe* = 0.010, *p*=.52) and transition type (*F*(1,15) = 1.65, *MSe* = 0.027, *p*=.22). A significant transition type x phonological similarity interaction emerged (*F*(1,15) = 11.39, *MSe* = 0.036, *p*=.004, *r*=.66), revealing a different pattern of learning and forgetting rates for similar and distinct word sequences. A simple main effects analysis reported a significantly higher learning rate for distinct compared to similar word sequences (*F*(1,15) = 11.35, *MSe* = 0.01, *p*=.004, *r*=.69). In contrast, a significantly higher forgetting rate was found for similar than distinct word sequences (*F*(1,15) = 7.87, *MSe* = 0.03, *p*=.013, *r*=.59).

For the nonword data, the ANOVA revealed a significant main effect of phonological similarity (*F*(1,15) = 10.95, *MSe* = 0.018, *p*=.005, *r*=.65), demonstrating higher overall transitional probabilities for similar nonword sequences. The main effect of transition type also attained significance (*F*(1,15) = 12.91, *MSe* = 0.098, *p*=.003, *r*=.68), confirming higher forgetting rates. A significant transition type x phonological similarity interaction emerged (*F*(1,15) = 5.89, *MSe* = 0.036, *p*=.028, *r*=.53), suggesting a differential pattern of learning and forgetting rates for similar and distinct nonword sequences. A simple main effects analysis revealed equivalent learning rates for similar and distinct nonword sequences (*F*(1.15) = 0.04, *MSe* = 0.01, *p*=.85), but a significantly higher forgetting rate for similar than distinct nonword sequences ($F(1,15) = 8.63$, $MSe = 0.05$, $p=.010$, $r=.60$).

One important consideration when applying the revised method of averaging over transitional steps concerns the distribution of missing data across these steps. It may be that a larger number of missing data are located at a specific transitional step(s). If this were the case, the procedure of averaging over transitional probabilities may suffer from bias. To address this issue, additional analyses were conducted on the word and nonword data whereby each missing transitional probability was replaced with that transitional step's mean transitional probability. If the data sets are free from bias due to the spread of missing data, these additional analyses should generate the same pattern of results as obtained when averaging over all transitional steps. Conducting this second analysis revealed an identical pattern of results for both word and nonword sequences, thereby confirming the absence of any bias in the spread of these missing values.

In summary, the results for word sequences revealed a higher learning rate for distinct sequences but a higher forgetting rate for similar sequences. The results for nonword sequences showed equivalent learning rates for similar and distinct sequences, but a higher forgetting rate for similar sequences.

5.2.2.3 Error Analysis

In line with Experiments 2 and 3, error analyses were conducted on the word and nonword data in order to further investigate the degree of stability of items within a sequence. Of issue was whether it is the case that once an item has been learned, it remains learned; and whether similar and distinct items display the same pattern. In particular, it was of interest to examine whether the slower learning of similar nonword sequences based on correct recall performance was due to fragility of nonwords.

Participants' responses were scored for a number of pre-defined error types: *total item errors*, *order errors* and *repetition errors*. Total item errors were defined as responses which differed from any of the items presented within a sequence (e.g. 'jug' as a response to the sequence 'nip, rod, wig, hem, tub, dot, yen'). Order errors were defined as instances in which an item was correctly recalled but in the incorrect serial position within a sequence (e.g. 'sem, seb, pez, *sef*, blank, blank' as a response to the sequence 'sem, seb, pez, pem, *sef*, pef). Repetition errors were defined as instances in which the same item was recalled at more than one serial position within a sequence (e.g. responding 'rag, cam, rap, *rat*, cab, ram, *rat*'). Finally, total item errors were further subdivided into: *omissions*, *intra-experimental intrusion (IEI) errors* and *extra-experimental intrusion (EEI) errors*. Omissions were defined as instances when participants failed to provide a response at one or more serial positions within a sequence (e.g. the response 'vek, lep, lef, vem, *blank*, *blank*'). IEI errors were defined as responses which differed from any of the items presented within a sequence but which appeared within the pool of items used within the experiment. EEI errors were defined as responses which differed from any of the items presented within a sequence and any of the items used within the experiment.

Proportions of each error type were calculated at each trial for F_1, F_2, and H for each of the four conditions. Mean proportional values were collapsed across materials A and B, and across F_1 and F_2. Mean proportional values were also averaged over every two trials. Proportional values were obtained by dividing the number of each error type by the number of items within a sequence. In order to "avoid confounding differences in order memory with different levels of item memory" (Allen & Hulme, 2006, p. 69), proportions of order errors per item recalled were calculated by dividing the total number of order errors by the number of items recalled regardless of order. Participants' data were excluded if an order error proportion could not be computed. Based on this criterion, no participants were excluded from the word conditions, but five participants were excluded from the nonword conditions. The same issue regarding the potential for bias due to the distribution of missing proportional values also applies here, as was the case for the Markov model analyses. Thus, additional analyses were also conducted on data sets replacing each missing order error proportion with that trial's mean order error proportion. Error types were analysed separately for word and nonword sequences.

The results of these error analyses revealed an extremely complex pattern of results which were difficult to comprehend and interpret. The complexity of these results failed to shed further insights into any interactive effects of phonological similarity and Hebb repetition on word and nonword sequence learning. Taking these points into consideration, the results of the majority of these error analyses are not reported here (but see Appendix 4(d) for tables showing the patterns of errors). However, the overall effects of phonological similarity and Hebb repetition on two main error types (total item errors and order errors) for word and nonword sequence learning will be reported.

Word Sequences: The pattern of errors shows that phonological similarity and Hebb repetition affected the propensity of errors (see Table 5.2 and Appendix 4(d)).

Table 5.2: *Mean proportions (and standard deviations) for the two main error types for Hebb and filler lists collapsed over trials for similar and distinct word sequences*

	List Type/Error Type			
	Filler		Hebb	
Condition	Item	Order	Item	Order
Similar Words	.454 (.096)	.470 (.149)	.379 (.131)	.338 (.156)
Distinct Words	.449 (.170)	.244 (.118)	.338 (.146)	.129 (.123)

Participants' error scores, expressed as proportions, were entered into two separate 2 x 2 repeated-measures ANOVAs incorporating two within-subjects factors: phonological similarity (similar items, distinct items) and list type (Hebb lists, filler lists).

For total item errors, the ANOVA revealed a significant main effect of list type ($F(1,15) = 16.49$, $MSe = 0.008$, $p=.001$, $r=.72$), demonstrating a higher proportion of total item errors for filler lists. The main effect of phonological similarity failed to reach significance ($F(1,15) = 0.66$, $MSe = 0.013$, $p=.43$) as did the phonological similarity x list type interaction ($F(1,15) = 1.01$, $MSe = 0.005$, $p=.33$).

For order errors, the ANOVA revealed significant main effects of phonological similarity ($F(1,15) = 66.27$, $MSe = 0.011$, $p<.0001$, $r=.90$) and list type ($F(1,15) = 12.91$, $MSe = 0.019$, $p=.003$, $r=.68$), indicating a higher proportion of order errors for similar sequences and for filler lists. The phonological similarity x list type interaction failed to attain significance (*F*(1,15) = 0.08, *MSe* = 0.015, *p*=.78). An identical pattern of results were obtained when conducting a second analysis in which missing order error proportions were replaced with trial means. This confirms the absence of any bias in the spread of these missing values.

Nonword Sequences: The pattern of errors shows that phonological similarity, and to a smaller extent Hebb repetition, affected the production of errors (see Table 5.3 and Appendix 4(d)).

Table 5.3: *Mean proportions (and standard deviations) for the two main error types for Hebb and filler lists collapsed over trials for similar and distinct nonword sequences*

	List Type/Error Type			
	Filler		Hebb	
Condition	Item	Order+	Item	Order+
Similar Nonwords	.465 (.067)	.331 (.142)	.496 (.146)	.475 (.219)
Distinct Nonwords	.633 (.101)	.109 (.066)	.622 (.196)	.103 (.117)

+ Based on 11 participants.

Participants' error scores, expressed as proportions, were entered into two separate 2 x 2 repeated-measures ANOVAs incorporating the same within-subjects factors as reported for the word sequences.

For total item errors, the ANOVA revealed a significant main effect of phonological similarity (*F*(1,15) = 17.67, *MSe* = 0.020, *p*=.001, *r*=.74), indicating a higher proportion of total item errors for distinct sequences. The main effect of list type failed to reach significance (*F*(1,15) = 0.19, *MSe* = 0.009, *p*=.67) as did the phonological similarity x list type interaction (*F*(1,15) = 0.43, *MSe* = 0.017, *p*=.52).

For order errors, the ANOVA revealed a significant main effect of phonological similarity (*F*(1,10) = 53.55, *MSe* = 0.018, *p*<.0001, *r*=.92), indicating a higher proportion of order errors for similar sequences. The main effect of list type failed to reach significance ($F(1,10) = 3.84$, $MSe = 0.014$, p=.08). The phonological similarity x list type interaction attained significance ($F(1,10) = 6.15$, $MSe = 0.010$, p=.003, r=.62), suggesting a differential effect of phonological similarity on order errors for Hebb and filler lists. A simple main effects analysis yielded a significantly higher proportion of order errors for similar than distinct sequences for both Hebb lists ($F(1,10) = 38.91$, $MSe = 0.01$, p=<.0001, r=.89) and filler lists ($F(1,10) = 35.98$, $MSe = 0.02$, $p<.0001$, r=.88). Thus, this interaction demonstrates a larger effect of phonological similarity on order errors for Hebb compared to filler lists. An identical pattern of results were obtained when conducting the analysis in which missing order error proportions were replaced with trial means, thus confirming the absence of any bias in the spread of these missing values.

5.2.3 Discussion

Experiment 4 represents the first in a series of experiments aimed at determining whether the Hebb repetition paradigm can be viewed as an analogue to the long-term learning of novel phonological word-forms. This was to be investigated by comparing patterns of results generated from the Hebb repetition task with those previously reported using the paired-associate task (e.g. Papagno & Vallar, 1992; Experiments 2 and 3). Thus, the aim of Experiment 4 was to investigate the effect of phonological similarity and Hebb repetition on the long-term learning of sequences of familiar and unfamiliar material. The main predictions were firstly that Hebb Effects would emerge for word and nonword sequences. Secondly, phonological similarity was expected to impair ISR performance for word and nonword sequences. Finally, it was predicted that phonological similarity would selectively disrupt the learning of nonword sequences, but not the learning of word sequences. The results of these predictions will be discussed for word and nonword sequences in turn.

Words

The analysis of correct recall performance provided evidence for a clear and reliable Hebb Effect for word sequences. Better overall performance was observed for Hebb compared to filler lists. Moreover, cumulative learning over trials for the Hebb sequence was confirmed by a list type x trials interaction. Although performance improved over trials for the filler lists, this is presumably attributable to non-specific practice effects, rather than the learning of these lists. The presence of a Hebb Effect demonstrates that participants can benefit from the repeated presentation of an ordered sequence of words. This finding replicates a number of studies reporting Hebb Effects for sequences of words (e.g. Cumming et al., 2006; Gagnon et al., 2004; Page et al., 2006; Sechler & Watkins, 1991).

The results also generated a pattern of findings in line with the prediction that phonological similarity would impair ISR performance but not the learning of word sequences. Firstly, a PSE was observed on overall ISR performance. Moreover, the size of this effect was comparable across Hebb and filler lists, as evidenced by a non-significant phonological similarity x list type interaction. Secondly, the absence of both a significant phonological similarity x trials interaction and a significant three-way interaction demonstrates that similar and distinct word sequences were learned at equivalent rates. Indeed, this was further supported by comparable performance increases over Trials 1 to 4 for similar (74% increase) and distinct (89% increase) repeated sequences.

The finding that phonological similarity impaired ISR performance suggests that participants are utilising phonological coding when recalling individual sequences. However, the learning of a repeated word sequence does not appear to rely on phonological coding, given the finding that phonological similarity failed to affect the learning of word sequences. To this end, it is proposed that the learning of word sequences does not utilise PSTM. This result appears to be consistent with the findings from paired-associate tasks (e.g. Papagno & Vallar, 1992; Experiments 2 and 3) and the predictions of the Burgess and Hitch (1999) model of the phonological loop.

In line with Experiment 2, an analysis of between-trial learning and forgetting rates was conducted using an application of a Markov model in order to investigate the degree of stability of repeated sequences. The results showed that phonological similarity affected both learning and forgetting rates for word sequences. Distinct sequences showed a significantly higher learning rate coupled with a significantly lower forgetting rate, demonstrating that distinct sequences were better learned and retained between trials than similar sequences. This suggests that once a distinct word has been learned in a particular position, it is more likely to remain learned between trials. In contrast, similar sequences appear more unstable, suggesting that a similar word learned in a particular position on one trial is more likely to be forgotten between trials.

The results based on between-trial learning and forgetting rates using the Markov model analysis might appear to contradict the results based on correct recall performance. Whereas the Markov model analysis showed better learning and less forgetting for distinct sequences, the correct recall performance analysis showed equivalent learning over trials for similar and distinct sequences. Learning based on correct recall performance tracks the number of items recalled in the correct serial position on each trial, but does not take into account whether the same items are recalled in the correct serial position from trial to trial. However, as was highlighted in Chapter Three, analysis of between-trial learning and forgetting rates takes both these points into account.

It is tentatively proposed that the observed differences in results between the Markov model and correct recall performance analyses may be related to two factors. The first of these refers to the idea that words learned at a particular position within a sequence are more likely to be retained between trials for distinct than similar sequences, as described above. The second factor refers to the better recall performance for distinct words on each individual trial. Taking both points into account, this means that there would be fewer words within a distinct sequence which still needed to be learned and recalled in the correct serial position on each trial compared to similar sequences. Thus, the transitional probability that the remaining unlearned distinct words would be learned on the next trial would be higher given the smaller pool of distinct words still left to be learned between trials. In contrast, given that fewer similar words are recalled on individual trials, the transitional probability that the remaining unlearned similar words would be learned on the next trial would be lower given the larger pool of similar words still left to be learned between trials. With this in mind, Tables 5.4 and 5.5 present a hypothetical example of this explanation. These show that despite equivalent learning rates for similar and distinct word sequences based on the correct recall performance results (see Table 5.4 and Figure 5.6), a higher between-trial learning rate is obtained for distinct than similar word sequences using a Markov model analysis (see Tables 5.5 and 5.6).

Table 5.4: *Data from a hypothetical participant for similar and distinct word sequences (√ = correct responses; x = incorrect responses)*

	Similar Word Sequences				Distinct Word Sequences			
Serial Position	Trial 1	Trial 2	Trial 3	Trial 4	Trial 1	Trial 2	Trial 3	Trial 4
1	√	x	√	√	√	√	√	√
2	x	√	√	√	√	x	x	√
3	x	√	x	√	√	√	√	√
4	x	√	√	x	x	√	√	√
5	x	x	√	√	x	√	√	√
6	x	x	x	x	x	√	√	√
7	x	x	x	√	x	x	√	√
Number Correct	1	3	4	5	3	5	6	7

Table 5.5: *Between-trial learning and forgetting rates using a Markov model based on the hypothetical participant's data shown in Table 5.4*

	Similar Word Sequences				Distinct Word Sequences			
Transitional Probability	T1-T2	T2-T3	T3-T4	*Mean*	T1-T2	T2-T3	T3-T4	*Mean*
Learning Rate	.50	.50	.67	.56	.75	.50	100	.75
Forgetting Rate	100	.33	.25	.53	.33	.0	.0	.11

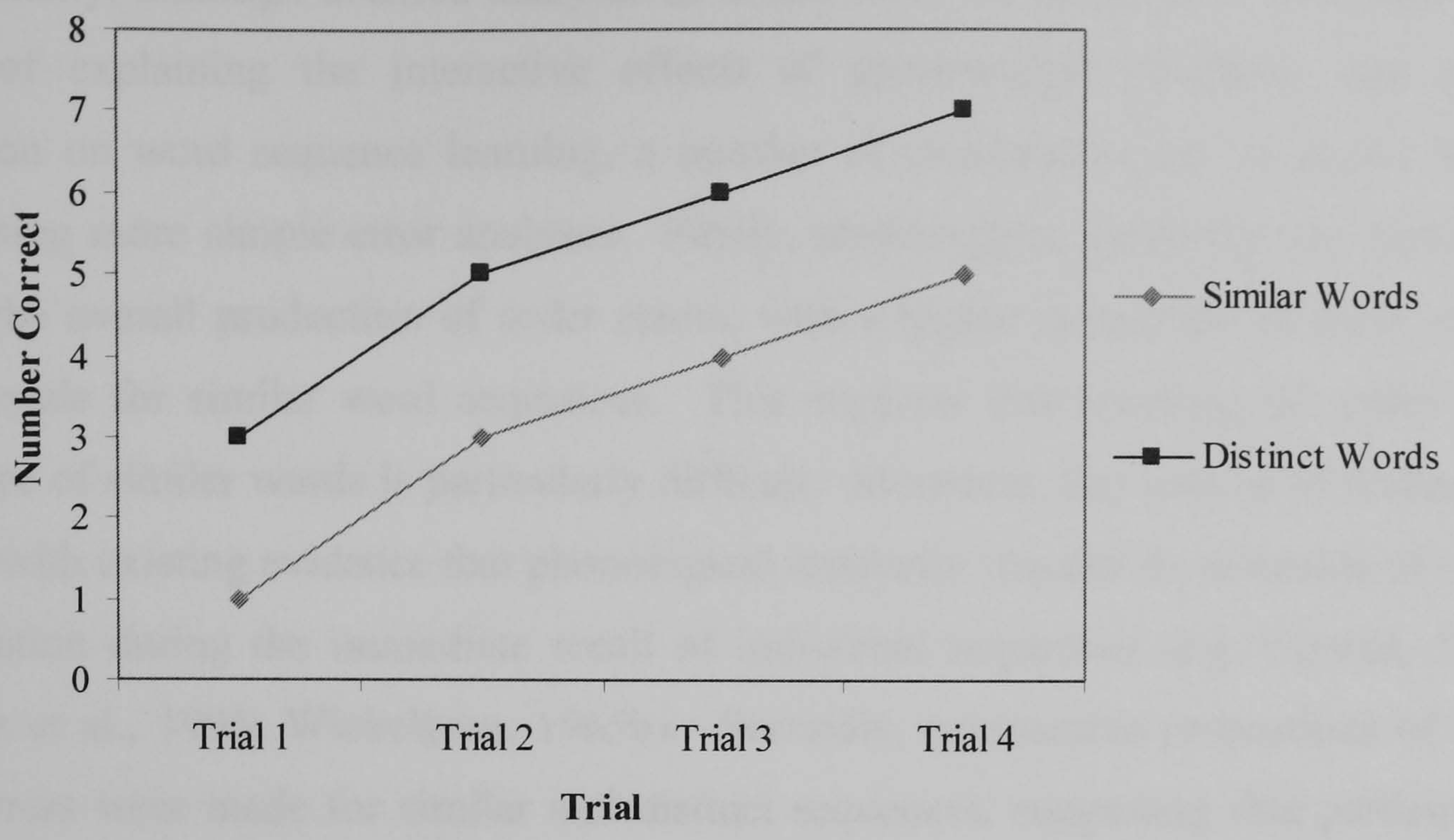


Figure 5.6: *Data from a hypothetical participant showing ISR performance for word sequences as a function of phonological similarity and trials*

Table 5.6: *Mean transitional probabilities for learning and forgetting transition types for similar and distinct word sequences based on the hypothetical participant's data in Figure 5.6*

	Condition	
Transition Type	Similar Word Sequences	Distinct Word Sequences
Learning	.56	.75
Forgetting	.53	.11

In sum, the finding of higher between-trial learning rates coupled with lower between-trial forgetting rates for word sequences in the Markov model analysis does not necessarily contradict the observation of equivalent learning over trials in the correct recall performance analysis. Rather, the results of the Markov model analysis reflect the differing amount of words to be learned between trials for similar and distinct word sequences. Moreover, the Markov model analysis has highlighted the fragility of similar word sequences; the higher between-trial forgetting rate observed for these sequences suggests that maintaining the order of similar words between trials is particularly difficult.

Finally, although detailed analyses of errors were not particularly informative in terms of explaining the interactive effects of phonological similarity and Hebb repetition on word sequence learning, a number of conclusions can be drawn when conducting more simple error analyses. Firstly, phonological similarity was shown to affect the overall production of order errors, with a higher proportion of these errors being made for similar word sequences. This suggests that recalling the order of a sequence of similar words is particularly difficult. Moreover, this pattern of findings is in line with existing evidence that phonological similarity impairs the retention of order information during the immediate recall of individual sequences (e.g. Conrad, 1965; Henson et al., 1996; Wickelgren, 1965b). Secondly, comparable proportions of total item errors were made for similar and distinct sequences, suggesting that participants are equally likely to recall a word not presented within a sequence for both types of sequence; this may reflect the fact that words are already familiar. Finally, in line with the emergence of a reliable Hebb Effect based on correct recall performance results, both total item errors and order errors were made more frequently for filler compared to Hebb lists. This suggests that Hebb repetition reduces the amount of errors generated.

Nonwords

The analysis of correct recall performance provided some evidence for a Hebb Effect for nonword sequences. Although the list type x trials interaction confirmed cumulative learning over trials for the Hebb sequences only, this was not supported by a main effect of list type. This suggests the presence of a somewhat weaker Hebb Effect than was obtained for word sequences. However, that cumulative learning was observed over trials suggests that participants do appear to benefit from the repeated presentation of a sequence of nonwords. This finding replicates previous studies that have reported Hebb Effects for sequences of nonwords (e.g. Gagnon et al., 2004; Turcotte et al., 2005).

The prediction that phonological similarity would impair ISR performance was not fully supported. No effect of phonological similarity was found on overall ISR performance. However, the significant three-way interaction showed a PSE on ISR performance for Hebb sequences, but not filler sequences. The finding that phonological similarity impaired ISR performance for Hebb sequences suggests that participants utilised phonological coding during the immediate recall of these sequences. In contrast, the lack of a PSE on ISR performance for filler sequences may suggest that participants abandoned the use of phonological coding when recalling these sequences (e.g. Baddeley, 2000b; Larsen & Baddeley, 2003). Alternatively, this finding may reflect that phonological similarity was ineffective given that only two nonwords were recalled in the correct serial position for similar and distinct filler sequences.

The prediction that phonological similarity would selectively disrupt the learning of nonword sequences was supported. The significant phonological similarity x trials interaction coupled with the significant three-way interaction showed learning over trials for the distinct Hebb sequences but no learning over trials for the similar Hebb sequences. This finding is further supported by a large difference in performance from Trials 1 to 4 for similar (15% increase) and distinct (115% increase) nonword sequences. No learning over trials was observed for either similar or distinct filler sequences. The finding that phonological similarity slowed down the learning of repeated similar nonword sequences suggests that, given the absence of lexical-semantic representations of nonwords in LTM, the learning of unfamiliar material relies upon phonological coding and as such is mediated by PSTM. These findings therefore converge with the nonword pair learning results of Papagno and Vallar (1992) and Experiments 2 and 3.

A further conclusion may be drawn from the nonword sequence learning results. The finding that a sequence of distinct nonwords can be learned over repeated presentations suggests that participants are capable of learning more than one novel phonological string of information (i.e. a nonword) simultaneously. This finding may somewhat support the natural process of vocabulary acquisition given that children are capable of learning several new words at once.

The analysis of between-trial learning and forgetting rates using a Markov model showed a higher forgetting rate for similar sequences but equivalent learning rates for similar and distinct sequences. This pattern of results suggests that the two types of sequence are equally difficult to learn but that similar sequences are more unstable and susceptible to forgetting between trials. As such, this pattern of findings may indicate that distinct nonwords learned in a particular position on one trial are often retained on the next trial. In contrast, similar nonwords that are learned in a particular position on one trial are more likely to be forgotten on the next trial. Interestingly, the current pattern of between-trial learning and forgetting rates obtained in the Markov model analysis parallels the corresponding results found for nonword pair learning in Experiment 2. Moreover, as was highlighted in Experiment 2, the present Markov model results are considered to inform the results based on correct recall performance by suggesting that similar nonwords are more vulnerable to forgetting between trials.

The finding that similar nonword sequences are more unstable between trials suggests that learning the order in which similar nonwords are presented within a sequence is particularly difficult. However, it may also be the case that this pattern of results indicates that phonological representations of similar nonwords are particularly fragile. As was suggested in Experiment 2, it may be that the phonemes within a similar nonword are not bound together very tightly and so subsequently fall apart between trials. Unfortunately, the Markov model analysis cannot distinguish between these two explanations; indeed, both of these explanations may contribute to the slower learning of similar nonword sequences over trials observed for correct recall performance.

Finally, as was found to be the case for word sequences, detailed analyses of errors revealed a complex pattern of results regarding the interactive effects of phonological similarity and Hebb repetition on nonword sequence learning. However, the results of a more simple series of error analyses revealed that phonological similarity impaired the immediate recall of a sequence of similar nonwords, as demonstrated by a higher proportion of order errors produced when recalling similar nonword sequences. This finding suggests that recalling the order of a sequence of similar nonwords is particularly difficult, and is line with the results of the Markov model analyses for nonword sequences reported above. Indeed, this pattern of results replicates previous studies that have found a detrimental effect of phonological similarity on retaining the order of a sequence of nonwords (e.g. Gathercole et al., 2001). Interestingly, more total item errors were made for distinct nonword sequences, suggesting that participants are more likely produce a nonword from outside the just-presented sequence when recalling sequences of distinct nonwords. Finally, Hebb repetition did not affect the production of either total item errors or order errors; this is in line with the somewhat weak Hebb Effect found for nonword sequences based on correct recall performance.

Limitations

Although the current findings offer support for the experimental predictions made, two limitations need to be acknowledged. The first of these refers to the relatively poor learning of nonword sequences. Learning of distinct nonword sequences only reached 2.8 nonwords (47%) by Trial 4. This contrasts with the learning of 5.3 words (76%) by Trial 4 for distinct word sequences. Moreover, robust learning of word-nonword pairs was observed in Experiments 2 and 3. The poor learning of nonword sequences observed in the current experiment may restrict the power of statistical analyses and may also, to some extent, limit any genuine effects of phonological similarity on learning. However, in contrast to Experiments 2 and 3, ceiling and floor effects were not observed in the current experiment.

The second limitation refers to problems encountered when confirming parallel patterns of results across materials A and B. Although the majority of these revealed converging patterns of results, a small number of interactions did emerge suggesting a differential pattern of results across the two sets of materials. These interactions may be a consequence of the experimental design adopted as, within each set of materials, all participants were presented with the same Hebb and filler sequences. As a result, the learning of a repeated sequence, and any effects of phonological similarity on this process, may be an artefact of the particular set of items comprising that sequence. Despite the fact that only a small number of participants (n=8) were presented with each materials set, it remains a possibility that the observed differences across materials may have contributed to the pattern of results obtained when conducting the main statistical analyses. To this end, future experiments should ensure that Hebb and filler sequences are rotated across participants in an attempt to eliminate the problems associated with confirming parallel patterns of results across more than one set of materials.

5.3 Chapter Summary

Experiment 4 generated a relatively clear pattern of results. Evidence was found for Hebb Effects for word and nonword sequences, although the Hebb Effect was somewhat weaker for nonword sequences. This suggests that participants can benefit from the repeated presentation of an ordered sequence of items. Phonological similarity was shown to affect overall ISR performance for word sequences, suggesting that participants utilise phonological coding during immediate recall of these sequences. In contrast, a PSE was observed on overall ISR performance for Hebb but not filler nonword sequences. Finally, phonological similarity failed to impair the rate of learning word sequences, confirming that the learning of familiar material does not rely on PSTM. In contrast, the learning of repeated nonword sequences was negatively affected by phonological similarity; slower learning of similar nonword sequences was observed. This provides further evidence that PSTM mediates the learning of unfamiliar material. These latter findings converge with those reported in paired-associate studies (e.g. Papagno & Vallar, 1992) and Experiments 2 and 3. The Markov model analysis of between-trial learning and forgetting rates indicated that phonological similarity had an effect on forgetting rates for both word and nonword sequences, suggesting that sequences of similar items are particularly unstable between trials. This conclusion is somewhat consistent with the findings from Experiment 2, which suggested that similar nonwords are fragile and are more likely to fall apart between trials. It is proposed that analyses based on a Markov model has offered additional insights into the learning of repeated sequences by highlighting the role of forgetting when learning such sequences. Finally, simple error analyses revealed that phonological similarity increased the propensity of order errors for both word and nonword sequences, suggesting that phonological similarity impairs the retention of order information.

In conclusion, it is proposed that Experiment 4 provides a relatively converging pattern of results across the two learning paradigms investigated in this thesis. Phonological similarity selectively disrupts the learning of unfamiliar material in both paired-associate and Hebb repetition tasks. Furthermore, phonological similarity fails to affect the learning of familiar material in the Hebb repetition task and, albeit to a lesser extent, the paired-associate task. Experiment 4 may therefore go some way to providing some support for the idea that the Hebb repetition task may possibly represent an analogue to new word-form learning. However, this experiment in isolation cannot provide conclusive evidence in support of this claim. Experiment 5 therefore aims to replicate Experiment 4 and also addresses its limitations.

Chapter Six: Replicating and Refining the Effects of Phonological Similarity and Hebb Repetition on Word and Nonword Sequence Learning

6.1 Introduction

Chapter Five generated initial evidence that the Hebb repetition paradigm may possibly represent an analogue to the long-term learning of novel phonological word-forms. Hebb Effects were found for both word and nonword sequences. Moreover, phonological similarity was shown to selectively disrupt the learning of nonword sequences but not the learning of word sequences, suggesting that PSTM is utilised when learning unfamiliar material but not when learning familiar material. These findings suggest a relatively converging pattern of results across the paired-associate and Hebb repetition tasks, providing further support for the PSTM hypothesis. The main aim of the current chapter therefore was to provide further evidence of convergence between paired-associate and Hebb repetition tasks.

6.2 Experiment 5: Effects of Phonological Similarity and Hebb Repetition on Sequence Learning with Reduced Demand on Item Learning

The aim of Experiment 5 was to provide a replication of Experiment 4. A second aim was to address the limitations observed in Experiment 4. In order to achieve this second aim, the design originally adopted in Experiment 4 was modified slightly. Given that the learning of nonword sequences was relatively poor in Experiment 4, particularly in comparison to the robust nonword pair learning observed in Experiments 2 and 3, it was decided to lower the task demands by reducing the number of unique items to learn across sequences.

In Experiment 4, as a result of presenting a different set of items for the Hebb list and each of the two filler lists, a total of 18 nonwords or 21 words had to be learned in each of the similar and distinct conditions. Presumably, 21 words will not pose huge difficulties for participants given that the words are already familiar. In this case, the task is to simply learn the order in which these words are presented in each sequence. In contrast, learning 18 nonwords per condition presumably places a huge demand on participants given that the items do not have existing lexical-semantic representations in LTM. As a result, participants not only have to learn the order in which the nonwords are presented within a sequence, but they also need to learn the nonwords themselves, that is, the order of the phonemes within each nonword. As such, participants need to conduct a two-stage learning process when learning sequences of nonwords, as opposed to a single-stage learning process for word sequences.

Given this particularly high demand on nonword learning, the current experiment attempted to lower this demand by presenting the same set of items for the Hebb list and each of the two filler lists. Participants would then be required to learn only 6 nonwords or 7 words in each of the similar and distinct conditions. Moreover, this design modification also serves to address the second limitation observed in Experiment 4 regarding the potential problems associated with confirming parallel patterns of results across two sets of materials. Presenting participants with the same set of items for each sequence eliminates the need for two separate sets of materials; each set of items used in Experiment 4 were presented to different participants in the current experiment. However, one disadvantage of this design is that it does not allow evidence to be obtained concerning the generalisability of results across more than one set of materials.

The same criteria were adopted for the detection of strong and weak Hebb Effects as in Experiment 4. To recap, a clear and reliable Hebb Effect would be indicated by a main effect of list type and a list type x trials interaction. Weaker evidence for a Hebb Effect would be indicated by a main effect of list type but no list type x trials interaction or vice versa. Poor evidence for a Hebb Effect would be indicated by a main effect of trials in the absence of a main effect of list type and a list type x trials interaction. Finally, the simultaneous failure to obtain a main effect of list type, a list type x trials interaction and a main effect of trials indicated the absence of a Hebb Effect.

The main predictions were the same as those made in Experiment 4. Firstly, it was predicted that Hebb Effects would emerge for both word and nonword sequences. Secondly, phonological similarity was predicted to impair ISR performance for both word and nonword sequences. Thirdly, it was hypothesised that phonological similarity would not affect the learning of word sequences. In contrast, it was predicted that phonological similarity would selectively disrupt the learning of nonword sequences, evidenced by slower learning of similar sequences. A novel prediction was also made. This predicted that lowering demand on item learning would improve sequence learning, particularly nonword sequence learning, compared with Experiment 4.

6.2.1 Method

6.2.1.1 Design

Participants undertook an ISR task as described in Experiment 4 with one critical exception: the Hebb list and each of the two filler lists comprised the same set of items. The Hebb list was presented every third trial interspersed with two non-repeated filler lists. Items in the Hebb list were presented in the same serial position on each trial and items in every filler list were presented in a different serial position on each trial. The experiment used a 2 x 2 x 8 x 2 mixed-design incorporating three within-subjects factors: phonological similarity (similar items, distinct items), list type (Hebb, filler) and trials (1 to 8); and one between-subjects factor: lexicality (words, nonwords). The phonological similarity and lexicality factors combined to produce four experimental conditions: similar word sequences; distinct word sequences; similar nonword sequences; and distinct nonword sequences.

Participants were randomly assigned to one of the lexicality conditions. Within each lexicality condition, the factor of phonological similarity was counterbalanced across participants so that half received similar sequences followed by distinct sequences, and half the reverse. For each experimental condition, participants were randomly assigned a single experimental set originally devised in Experiment 4. This set served as the Hebb list (H) and each of the two filler lists (Filler 1, F_1; Filler 2, F_2). The presentation order of the list type factor was F_1, followed by F_2, followed by H. For each experimental set within each experimental condition, the order of items within a list remained the same across participants.

6.2.1.2 Participants

Forty-eight individuals volunteered to participate for either course credit or payment. Forty-five were students (undergraduate and postgraduate) and staff attending the University of York. Three were sixth-form students attending Bootham School in York. There were 34 females and 14 males aged between 17 years and 46 years (mean age of 22.3 years). All participants spoke English as their native language and reported no known hearing or language impairments.

6.2.1.3 Apparatus

The same apparatus were used as in Experiment 4.

6.2.1.4 Materials

The materials were the same as those used in Experiment 4. Each word and nonword set remained identical to that devised in Experiment 4.

6.2.1.5 Procedure

The procedure was identical to that described in Experiment 4. Responses were scored both immediately using a strict serial recall criterion and later using the minidisc recordings to verify scoring and to transcribe the exact pronunciation of each item. Maximum scores for each list were 7 for word sequences and 6 for nonword sequences.

6.2.2 Results

6.2.2.1 Correct Recall Analysis

Responses were scored following the same strict serial recall criterion as described in Experiment 1. No instances occurred in which a participant consistently made the same mispronunciation of an item. Instances in which the correct pronunciation of an item could not be clearly determined were scored as correct, provided they were recalled in the correct serial position; this occurred on five occasions[38].

---

[38] The word '*nut*' was perceived as '*knot*', presumably due to accent differences between the participant and experimenter. The nonword '*kem*' was difficult to determine from '*ken*'; the nonword '*jal*' was perceived as '*gel*'; the nonword '*som*' was difficult to determine from '*zom*'. The pronunciation of the vowel /o/ by one participant was perceived by the experimenter as the vowel /u/ (e.g. the nonword '*yom*' was perceived as '*yum*'), presumably due to accent differences between the participant and experimenter.

Data from two participants proved to be outliers: one participant from the word condition and one participant from the nonword condition. These participants' data were excluded from all further statistical analyses[39]. The data were normally distributed following the exclusion of these data.

The means and standard deviations were calculated for F_1, F_2, and H based on the 23 participants in each lexicality condition. As in Experiment 4, mean scores were averaged over pairs of trials. In order to confirm parallel patterns of results across F_1 and F_2, the data from each lexicality condition were subjected to separate repeated-measures ANOVAs. Parallel patterns of results were found across F_1 and F_2 for both the word and nonword data (all *p*s<.05), thus allowing these two list types to be collapsed for all further statistical analyses.

The word and nonword data were analysed separately, as in Experiment 4. Participants' scores were entered into two separate three-way repeated measures ANOVAs each incorporating three within-subjects factors: phonological similarity (similar items, distinct items), list type (Hebb, filler) and trials (1 to 4).

The ANOVA conducted on the word data reported significant main effects of phonological similarity ($F(1,22) = 33.39$, $MSe = 4.437$, $p<.0001$, $r=.78$) and list type ($F(1,22) = 7.84$, $Mse = 2.327$, $p=.010$, $r=.51$), confirming better recall performance for distinct word sequences, and for Hebb lists. The main effect of trials also attained significance ($F(3,66) = 7.34$, $Mse = 0.586$, $p<.0001$, $\eta_p^2=.25$), demonstrating learning over trials (see Figure 6.1). All interactions failed to reach significance (all *p*s>.05).

---

[39] Outliers were defined as in Experiment 1 and were calculated as described in Experiment 4.

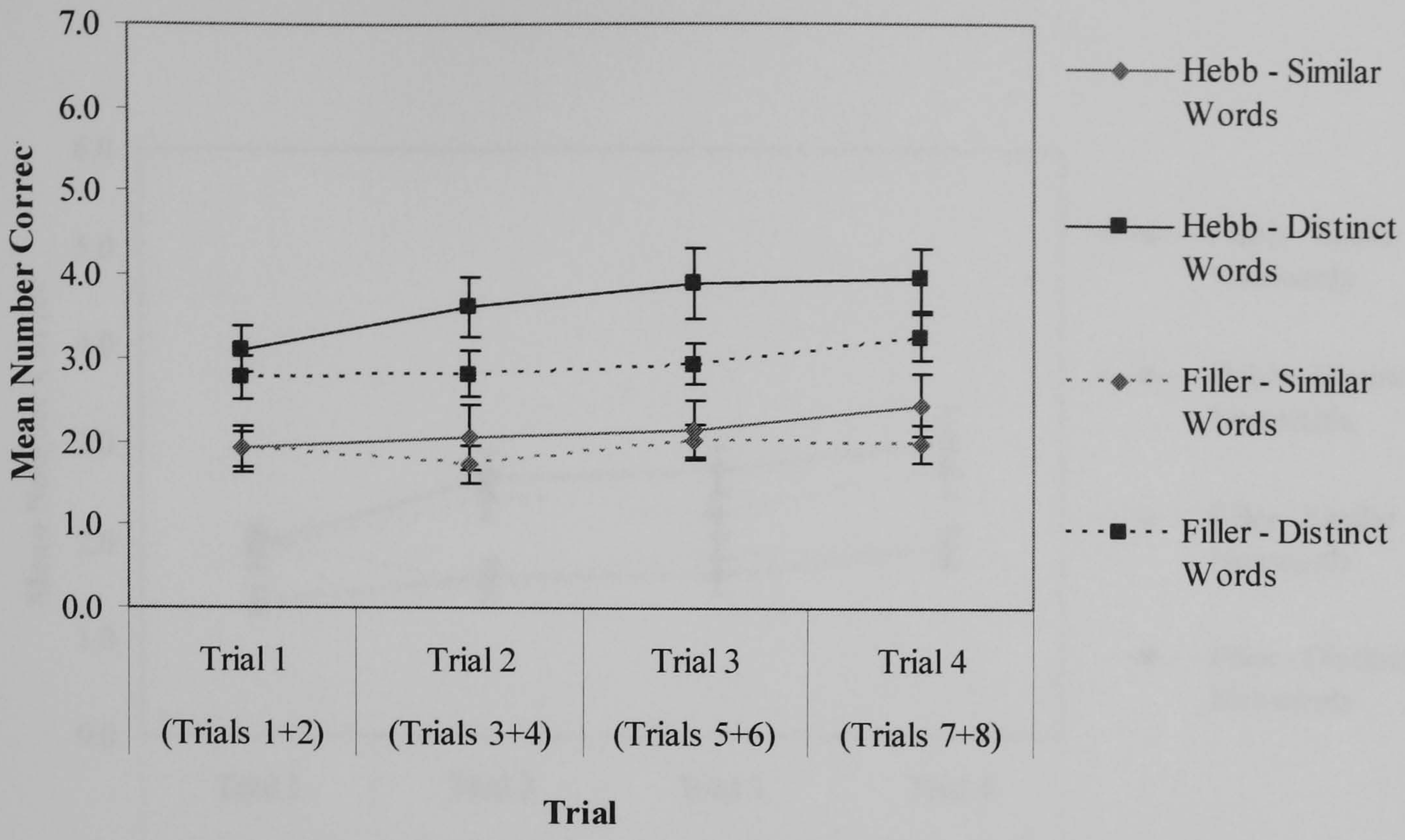


Figure 6.1: *ISR performance for word conditions as a function of phonological similarity, list type and trials (error bars represent standard error of the mean)*

The ANOVA conducted on the nonword data reported a non-significant main effect of list type ($F(1,22) = 0.01$, $Mse = 0.655$, $p=.92$). The main effect of phonological similarity attained significance ($F(1,22) = 17.68$, $Mse = 2.670$, $p<.0001$, $r=.67$) as did the main effect of trials ($F(3,66) = 9.07$, $Mse = 0.633$, $p<.0001$, $\eta_p^2=.29$), demonstrating better performance for distinct nonword sequences, and learning over trials (see Figure 6.2).

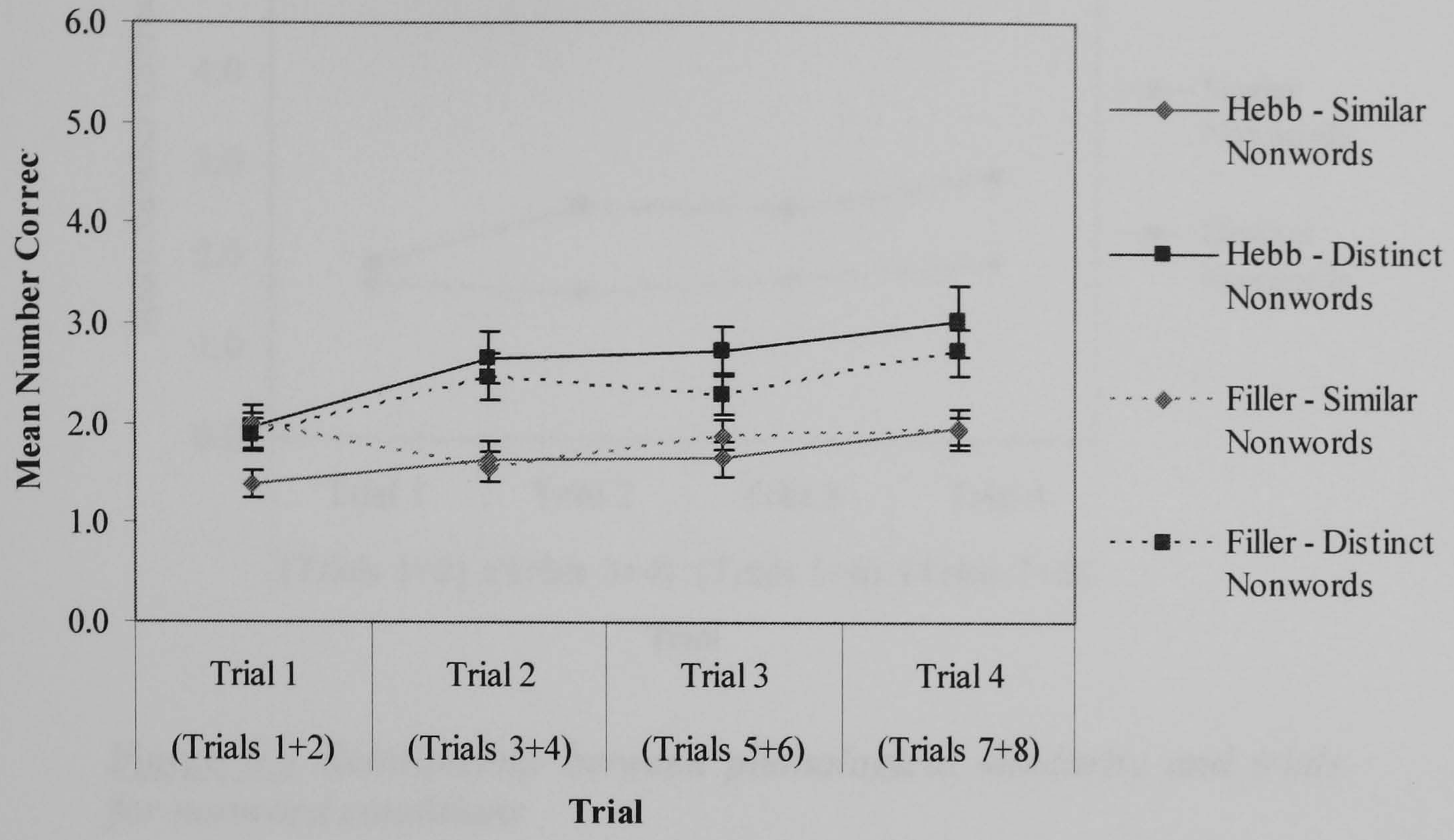


Figure 6.2: *ISR performance for nonword conditions as a function of phonological similarity, list type and trials (error bars represent standard error of the mean)*

The phonological similarity x trials interaction attained significance (*F*(3,66) = 4.67, *Mse* = 0.542, *p*=.005, η_p^2=.18), suggesting a differential pattern of learning over trials for similar and distinct nonword sequences (see Figure 6.3). A simple main effects analysis revealed significant learning over trials for distinct nonword sequences (*F*(3,66) = 12.44, *Mse* = 0.58, *p*<.0001) but not for similar nonword sequences (*F*(3,66) = 1.76, *Mse* = 0.59, *p*=.16).

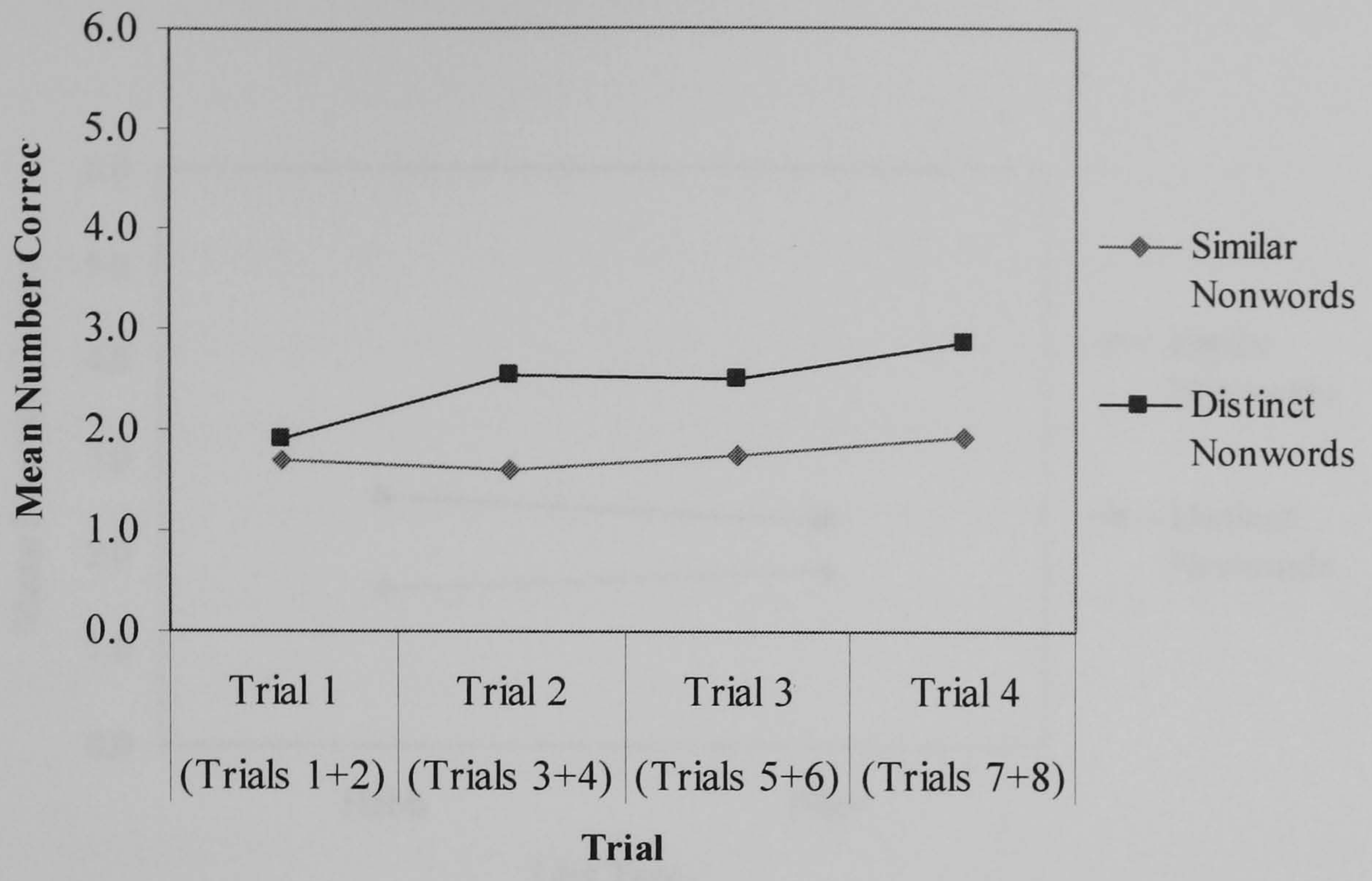


Figure 6.3 *Relationship between phonological similarity and trials for nonword conditions*

A marginally non-significant phonological similarity x list type interaction emerged from the analysis (*F*(1,22) = 4.24, *Mse* = 1.036, *p*=.052, *r*=.40), indicating differential effects of phonological similarity for Hebb and filler lists (see Figure 6.4). Given that this interaction almost attained significance, it was subjected to a simple main effects analysis. This confirmed the presence of significant PSEs for both Hebb lists (*F*(1,22) = 15.04, *Mse* = 2.67, *p*=.001, *r*=.64) and filler lists (*F*(1,22) = 11.03, *Mse* = 1.03, *p*=.003, *r*=.58). Thus, this interaction demonstrates a larger PSE for Hebb compared to filler lists. The list type x trials interaction failed to reach significance (*F*(3,66) = 1.49, *Mse* = 0.588, *p*=.23) as did the three-way interaction (*F*(3,66) = 0.88, *Mse* = 0.504, *p*=.46).

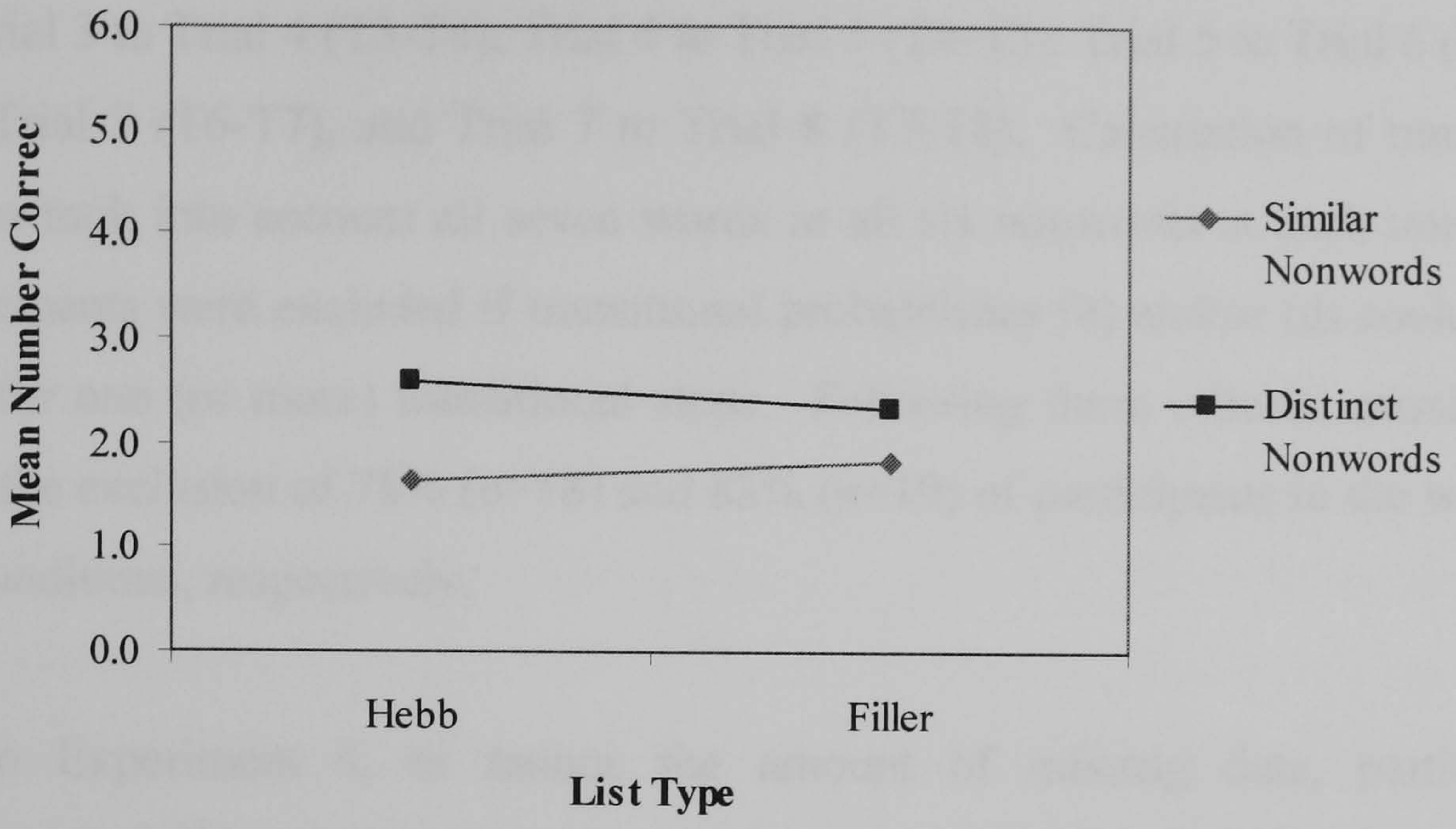


Figure 6.4 *Relationship between phonological similarity and list type for nonword conditions*

In summary, the results for word sequences showed weak evidence for a Hebb Effect. A clear PSE on ISR performance was observed. Similar and distinct repeated word sequences revealed equivalent rates of learning over trials. The results for nonword sequences provided poor evidence for a Hebb Effect. However, a clear PSE was observed on ISR performance. Faster learning over trials was observed for distinct repeated sequences compared to similar sequences.

6.2.2.2 Markov Model Analysis

A comparison of learning and forgetting rates was conducted using an application of a two-state Markov model in order to further investigate the degree of stability of items within a repeated sequence and also to form a comparison with Experiment 4.

Learning and forgetting rates were calculated for the 23 participants in each of the two lexicality conditions following the procedure described in Experiment 4. Learning rates represent the transitional probability of an item at a particular position within a sequence that was in an unlearned state on one trial, being learned on the next trial (b). Forgetting rates represent the transitional probability of an item at a particular position within a sequence that was in a learned state on one trial, being forgotten on the next trial (d).

For each participant, transitional probabilities were calculated for each of seven transitional steps for the Hebb lists only: Trial 1 to Trial 2 (T1-T2), Trial 2 to Trial 3 (T2-T3), Trial 3 to Trial 4 (T3-T4), Trial 4 to Trial 5 (T4-T5), Trial 5 to Trial 6 (T5-T6), Trial 6 to Trial 7 (T6-T7), and Trial 7 to Trial 8 (T7-T8). Calculation of transitional probabilities took into account all seven words or all six nonwords at each transitional step. Participants were excluded if transitional probabilities (b) and/or (d) could not be calculated for one (or more) transitional steps. Following these criteria, missing data resulted in the exclusion of 78% (n=18) and 83% (n=19) of participants in the word and nonword conditions, respectively.

As in Experiment 4, to reduce the amount of missing data, participants' transitional probabilities were averaged over all seven transitional steps. Participants' data were excluded if an averaged transitional probability could not be calculated due to missing data at all seven transitional steps in any one of the two conditions. No participants were excluded following this revised criterion. Averaged transitional probabilities were therefore calculated based on all 23 participants in each lexicality condition.

Similar patterns of results were observed for words and nonwords, with phonological similarity appearing to affect both learning and forgetting rates (see Table 6.1). Higher learning rates and lower forgetting rates were shown for distinct compared with similar sequences, for both word and nonword conditions.

Table 6.1: *Mean (and standard deviation) transitional probabilities for learning and forgetting transition types for similar and distinct word and nonword sequences*

	Condition			
	Word Sequences		Nonword Sequences	
Transition Type	Similar	Distinct	Similar	Distinct
Learning	.239 (.198)	.338 (.266)	.181 (.064)	.268 (.140)
Forgetting	.542 (.290)	.249 (.228)	.477 (.232)	.301 (.162)

In line with analyses based on correct recall performance, the word and nonword data were analysed separately. Participants' averaged transitional probability values were entered into two-way repeated measures ANOVAs each incorporating two within-subjects factors: transition type (learning, forgetting) and phonological similarity (similar, distinct).

For the word data, the ANOVA revealed a significant main effect of phonological similarity ($F(1,22) = 6.99$, $Mse = 0.031$, $p=.015$, $r=.49$), demonstrating higher overall transitional probabilities for similar word sequences. The main effect of transition type failed to attain significance ($F(1,22) = 2.03$, $Mse = 0.131$, $p=.17$). A significant transition type x phonological similarity interaction emerged ($F(1,22) = 17.64$, $Mse = 0.050$, $p<.0001$, $r=.67$), suggesting a differential pattern of learning and forgetting rates for similar and distinct word sequences. A simple main effects analysis revealed statistically comparable learning rates for similar and distinct word sequences ($F(1,22) = 3.81$, $Mse = 0.03$, $p=.064$) but a significantly higher forgetting rate for similar word sequences ($F(1,22) = 19.02$, $Mse = 0.05$, $p<.0001$, $r=.68$).

The ANOVA conducted on the nonword data reported a significant main effect of transition type ($F(1,22) = 18.72$, $MSe = 0.033$, $p<.0001$, $r=.68$), demonstrating higher overall forgetting rates. The main effect of phonological similarity failed to reach significance ($F(1,22) = 1.77$, $Mse = 0.025$, $p=.20$). The transition type x phonological similarity interaction attained significance ($F(1,22) = 16.79$, $Mse = 0.024$, $p<.0001$, $r=.66$), indicating a differential pattern of learning and forgetting rates for similar and distinct nonword sequences. A simple main effects analysis revealed a significantly higher learning rate for distinct nonword sequences ($F(1,22) = 6.07$, $Mse = 0.01$, $p=.022$, $r=.47$) but a significantly higher forgetting rate for similar nonword sequences ($F(1,22) = 10.29$, $Mse = 0.03$, $p=.004$, $r=.56$).

As in Experiment 4, further analyses were conducted on the word and nonword data in order to confirm that the data sets were free from bias due to the spread of missing data across transitional steps. Missing transitional probabilities were replaced with transitional step means.

For the word sequences, this second analysis revealed the same pattern of results as the initial analysis. However, the main effect of transition type attained significance (*F*(1,22) = 5.27, *Mse* = 0.049, *p*=.032, *r*=.44), demonstrating higher overall forgetting rates. All remaining statistical results were identical to the initial analysis. For nonword sequences, the second analysis revealed an identical pattern of results to the initial analysis. These analyses were therefore considered to confirm the absence of any bias in the spread of these missing values.

In summary, the word sequence results revealed comparable learning rates for similar and distinct sequences and a higher forgetting rate for similar sequences. The results for the nonword sequences showed a higher learning rate for distinct sequences and a higher forgetting rate for similar sequences.

6.2.2.3 Error Analyses

Error analyses were conducted on the word and nonword data to further investigate the degree of stability of items within a sequence and to form a comparison with Experiment 4. However, as was found in Experiment 4, these analyses produced a complex pattern of results which were difficult to understand and subsequently interpret in terms of the interactive effects of Hebb repetition and phonological similarity on word and nonword sequence learning. The majority of these analyses are therefore not reported here (but see Appendix 5(a) for tables showing the patterns of errors). However, in line with Experiment 4, simple error analyses were conducted on the two main error types (total item errors and order errors) in order to determine the overall effect of phonological similarity and Hebb repetition on word and nonword sequence learning.

Participants' responses were scored and calculated for each of the two main error types following the same criteria as described in Experiment 4. Mean proportional values were collapsed across F_1 and F_2 and were averaged over every two trials. Order errors were calculated following the revised procedure: the total number of order errors was divided by the number of items recalled regardless of order. When computing the revised order error proportions, data from two participants from the nonword conditions but no participants from the word conditions were excluded. As in Experiment 4, analyses were conducted on word and nonword data separately.

Word Sequences: The pattern of errors shows that phonological similarity and Hebb repetition affected the production of errors (see Table 6.2 and Appendix 5(a)).

Table 6.2: *Mean proportions (and standard deviations) for the two main error types for Hebb and filler lists collapsed over trials for similar and distinct word sequences*

	List Type/Error Type			
	Filler		Hebb	
Condition	Item	Order	Item	Order
Similar Words	.413 (.123)	.509 (.162)	.392 (.158)	.511 (.210)
Distinct Words	.339 (.127)	.338 (.161)	.276 (.135)	.274 (.238)

Participants' error scores, expressed as proportions, were entered into two separate two-way repeated-measures ANOVAs incorporating two within-subjects factors: phonological similarity (similar items, distinct items) and list type (Hebb lists, filler lists).

For total item errors, the ANOVA revealed significant main effects of phonological similarity ($F(1,22) = 12.87$, $MSe = 0.016$, $p=.002$, $r=.61$) and list type ($F(1,22) = 10.17$, $MSe = 0.004$, $p=.004$, $r=.56$), demonstrating a higher proportion of total item errors for similar sequences and for filler lists. The phonological similarity x list type interaction failed to reach significance ($F(1,22) = 1.97$, $MSe = 0.005$, $p=.18$).

For order errors, the ANOVA revealed a significant main effect of phonological similarity ($F(1,22) = 49.73$, $MSe = 0.019$, $p<.0001$, $r=.83$), indicating a higher proportion of order errors for similar sequences. The main effect of list type failed to attain significance ($F(1,22) = 1.78$, $MSe = 0.013$, $p=.20$) as did the phonological similarity x list type interaction ($F(1,22) = 2.04$, $MSe = 0.013$, $p=.17$). As in Experiment 4, a second analysis was conducted on the order error data in order to confirm the absence of bias due to the spread of missing order error proportions. This analysis revealed an identical pattern of results as the initial analysis.

Nonword Sequences: The pattern of errors reveals that phonological similarity and Hebb repetition affected the production of errors (see Table 6.3 and Appendix 5(a)).

Table 6.3: *Mean proportions (and standard deviations) for the two main error types for Hebb and filler lists collapsed over trials for similar and distinct nonword sequences*

	List Type/Error Type			
	Filler		Hebb	
Condition	Item	Order+	Item	Order+
Similar Nonwords	.441 (.098)	.399 (.107)	.457 (.091)	.469 (.159)
Distinct Nonwords	.522 (.131)	.178 (.116)	.457 (.169)	.206 (.149)

$^{+}$ Based on 20 participants.

Participants' error scores, expressed as proportions, were entered into two separate two-way repeated-measures ANOVAs incorporating the same within-subjects factors as reported for the word sequences.

For total item errors, the ANOVA reported non-significant main effects of phonological similarity (F(1,22) = 1.85, MSe = 0.021, p=.19) and list type (F(1,22) = 2.71, MSe = 0.005, p=.11). The phonological similarity x list type interaction attained significance (F(1,22) = 8.26, MSe = 0.004, p=.009, r=.52), demonstrating a different effect of phonological similarity on total item errors for Hebb and filler lists. A simple main effects analysis yielded a significantly higher proportion of total item errors for similar than distinct sequences for filler lists (F(1,22) = 0.00, MSe = 0.02, p=.98), but comparable proportions of total item errors for similar and distinct sequences for Hebb lists (F(1,22) = 9.56, MSe = 0.01, p=.005, r=.55).

For order errors, the ANOVA revealed a significant main effect of phonological similarity (F(1,20) = 62.63, MSe = 0.020, p>.0001, r=.87), demonstrating a higher proportion of order errors for similar sequences. The main effect of list type failed to attain significance (F(1,20) = 3.13, MSe = 0.016, p=.09) as did the interaction (F(1,20) = 0.55, MSe = 0.017, p=.47). An identical pattern of results were obtained when conducting the analysis whereby missing order error proportions were replaced with trial means, thus confirming the absence of any bias in the spread of these missing values.

6.2.3 Discussion

The primary aim of Experiment 5 was to replicate the results of Experiment 4. The second aim was to address the main limitation observed in Experiment 4 regarding the relatively poor learning of nonword sequences. It was hypothesised that this limitation may be due to the large number of unique nonwords participants were required to learn as a result of presenting a different set of items for the Hebb list and each of the two filler lists. Thus, in an attempt to lower this demand on nonword learning, the current experiment presented the same set of items for the Hebb list and each of the two filler lists. The main predictions made were the same as in Experiment 4. One novel prediction was also made. This predicted that lowering demand on item learning would improve sequence learning, particularly nonword sequence learning, compared to Experiment 4. The results of each prediction will be discussed for word and nonword sequences in turn.

Words

The analysis of correct recall performance provided some support for the original prediction that a Hebb Effect would emerge for word sequences. Although better overall performance was observed for Hebb than filler sequences, this was not supported by a significant list type x trials interaction. However, the main effect of trials demonstrated that learning did occur. As such, this pattern of results suggests that cumulative learning over trials was observed for the repeated sequences but that this learning was not significantly better than the learning observed over trials for the filler sequences. This finding differs somewhat from Experiment 4 in which cumulative learning over trials was greater for the Hebb sequences than for the filler sequences. However, the current results do suggest that participants can benefit from the repeated presentation of an ordered sequence of words.

The results generated a pattern of findings in line with the original prediction that phonological similarity would impair ISR performance. A PSE was observed on overall ISR performance. Moreover, the size of the PSE was comparable for Hebb and filler lists. These findings are consistent with the results of Experiment 4 and provide further evidence that participants rely on phonological coding during the immediate recall of word sequences.

The original prediction that phonological similarity would not affect the learning of repeated word sequences was supported. Equivalent rates of learning similar and distinct word sequences were confirmed by the lack of both a significant phonological similarity x trials interaction and a significant three-way interaction. Moreover, this result is further supported by the observation of comparable increases in performance over Trials 1 to 4 for similar (36% increase) and distinct (29% increase) repeated sequences. However, it is worth noting that these increases in performance are considerably lower that those reported in Experiment 4 (74% and 89% increase for similar and distinct word sequences, respectively). Indeed, this reflects the weaker Hebb Effect observed for word sequences in the current experiment compared to Experiment 4.

The finding that phonological similarity did not impair the learning of repeated word sequences suggests that participants were not utilising phonological coding when learning these sequences. This parallels the results obtained in Experiment 4 and provides further evidence that participants rely on existing lexical and semantic representations of words in LTM in order to learn a repeated sequence of words. Moreover, the present findings also converge with the results of previous studies (e.g. Fallon et al., 2005; Hitch et al., in press; Page et al., 2006) and are in line with the predictions of the Burgess and Hitch (1999) model of the phonological loop. Importantly, the current findings also appear to be consistent with the results of Papagno and Vallar (1992), in which phonological similarity failed to impair the learning of word-word pairs, and thereby provide further support for the claim that PSTM does not mediate the learning of familiar material.

In line with Experiment 4, an analysis of between-trial learning and forgetting rates was conducted on the repeated word sequences using an application of a Markov model. Phonological similarity was shown to have a negative impact on between-trial forgetting rates only; comparable between-trial learning rates were found for similar and distinct sequences. This seems to suggest that although the two types of word sequences are equally difficult to learn, similar word sequences are particularly unstable and likely to be forgotten from trial to trial. However, this pattern of between-trial learning and forgetting rates differs somewhat from the corresponding results of Experiment 4. Although a higher forgetting rate was also shown for similar than distinct word sequences in Experiment 4, thereby confirming the unstable nature of similar word sequences between trials, a higher learning rate was found for distinct than similar sequences in Experiment 4. It is suggested that this difference in the pattern of between-trial learning rates is related to the poorer learning of distinct and similar repeated word sequences based on correct recall performance in the current experiment, as previously highlighted. As a result of this poor learning, the Markov model analyses are considered of limited value in the present experiment.

Finally, detailed error analyses revealed a complex pattern of results which were difficult to interpret. However, the results of simple error analyses showed that phonological similarity impaired the immediate recall of a sequence of similar words, as demonstrated by a higher proportion of order errors produced for similar sequences. This suggests that sequences of similar words are particularly difficult to recall. This finding is in line with the results of Experiment 4. Moreover, this result converges with previous studies which have reported negative effects of phonological similarity on learning the order of a sequence of words (e.g. Conrad, 1965; Gathercole et al., 2001; Henson et al., 1996; Wickelgren, 1965b). Phonological similarity was also shown to increase the propensity of total item errors, indicating that participants are more likely to erroneously recall a word from outside a just-presented sequence of similar words. Hebb repetition affected the production of total item errors only; a higher proportion of these errors were made for filler compared to Hebb lists. The finding that Hebb repetition only affected total item errors and not order errors may be related to the weak evidence of a Hebb Effect for word sequences based on correct recall performance.

Nonwords

The analysis of correct recall performance provided only minimal support for the original prediction that a Hebb Effect would emerge for nonword sequences. Firstly, comparable levels of overall performance were observed for Hebb and filler lists. Secondly, although the main effect of trials confirmed that there was cumulative learning over trials, the absence of a significant list type x trials interaction suggests that learning over trials for Hebb sequences did not exceed learning over trials for filler sequences. This finding is in contrast to Experiment 4 which showed better learning over trials for Hebb than filler sequences. The current results therefore provide only poor evidence for a Hebb Effect.

The original prediction that phonological similarity would impair ISR performance for nonword sequences was supported by the presence of a significant PSE on overall ISR performance. Moreover, whereas a PSE was only found for Hebb sequences in Experiment 4, significant PSEs were observed for both Hebb and filler sequences in the current experiment. This provides evidence that participants utilise phonological coding within PSTM during the immediate recall of both repeated and filler nonword sequences. It was suggested in Experiment 4 that the lack of a PSE on ISR performance for filler nonword sequences may reflect that participants either abandoned the use of phonological coding during the immediate recall of these sequences or that too few nonwords were recalled for an effect of phonological similarity to emerge. The present findings suggest that the latter explanation may be appropriate given that participants recalled a larger number of nonwords for filler nonword sequences in the current experiment. Indeed, this may be due, in part, to the fact that participants were presented with the same set of nonwords for the Hebb and filler sequences.

The original prediction that phonological similarity would disrupt the learning of nonword sequences was difficult to reliably assess given the poor evidence for a Hebb Effect for these sequences. Although the significant phonological similarity x trials interaction demonstrated learning over trials for distinct sequences but not for similar sequences, the absence of a significant three-way interaction suggests that this pattern of results is the same for both Hebb and filler lists. That learning occurred for the distinct filler sequences may not be entirely surprising; some item learning may be expected over repeated presentations despite the lack of consistent order information. However, despite the statistical pattern of results, visual inspection of Figure 6.2 (p. 218) seems to suggest comparable rates of learning the similar and distinct repeated sequences.

Indeed comparable increases in performance over Trials 1 to 4 were observed for similar (43% increase) and distinct (50% increase) repeated sequences. On the basis of these results, it is difficult to clearly determine whether participants utilise PSTM when learning sequences of unfamiliar material due to the absence of a reliable Hebb Effect.

The Markov model analysis of between-trial learning and forgetting rates conducted on the repeated nonword sequences revealed that phonological similarity affected both learning and forgetting rates. Similar nonword sequences showed a lower learning rate coupled with a higher forgetting rate. This suggests that not only are similar nonword sequences more difficult to learn than distinct sequences, they are also more susceptible to forgetting between trials. However, as was found for the word sequences, the current pattern of between-trial learning and forgetting rates is inconsistent with the corresponding results from Experiment 4. Although similar nonword sequences showed a higher forgetting rate, confirming that similar sequences are particularly fragile, equivalent learning rates were observed for similar and distinct sequences in Experiment 4. As was suggested for the word sequences, this discrepancy is presumably related to the poor learning of nonword sequences based on correct recall performance in the current experiment. Indeed, this idea is further substantiated by poor evidence of a Hebb Effect for nonword sequences in the current experiment. To this end, it is proposed that any interpretation of the current Markov model results is limited.

Finally, detailed error analyses failed to provide further insights into the interactive effects of phonological similarity and Hebb repetition on nonword sequence learning. However, in line with Experiment 4 and previous studies (e.g. Gathercole et al., 2001), simple error analyses revealed an overall effect of phonological similarity on recalling the order of a sequence of nonwords; more order errors were made for similar compared to distinct sequences. This further confirms that recalling a sequence of nonwords in the correct order is more difficult when that sequence contains similar nonwords as opposed to distinct nonwords. Phonological similarity was also shown to affect the production of total item errors: a higher proportion of these errors were made for similar than distinct nonword sequences. However, this was found to be the case for filler lists but not for Hebb lists, as indicated by the phonological similarity x list type interaction. This suggests that participants more often recall a nonword from outside the pool of nonwords within the just-presented sequence for similar than distinct nonword sequences. This finding is in contrast to the corresponding results of Experiment 4, which showed a higher proportion of total item errors for distinct nonword sequences for both Hebb and filler lists. Hebb repetition did not affect the production of either total item errors or order errors. This may reflect the current experiment's lack of a reliable Hebb Effect for nonword sequences based on correct recall performance.

Hebb Effect Issues

It is important to consider why the current experiment found weaker Hebb Effects for both word and nonword sequences in comparison to Experiment 4. The main change between the two experiments was to present the same set of items for the Hebb and filler lists in the current experiment. This suggests that the current weak Hebb Effects may be due to some form of interference between the Hebb and filler lists. It may be that the filler sequences interfere with the learning of repeated sequences. For example, presenting the same set of items in the Hebb and filler lists may facilitate partial learning of the filler sequences, which may subsequently interfere with learning the repeated sequences. In turn, this may serve to reduce the likelihood of obtaining a reliable Hebb Effect for repeated sequences.

On this basis, it is tentatively suggested that the Hebb Effect is influenced by the degree of item overlap between Hebb and filler lists. Reliable Hebb Effects may be more likely to emerge when there is less opportunity for interference between these two list types, such as when different sets of items are presented for Hebb and filler lists. Indeed, the present findings are in line with some recent data. Cumming et al. (2006) failed to find evidence of reliable sequence learning for word sequences when the items in the Hebb and filler lists overlapped. The authors concluded that the critical factor in generating reliable Hebb Effects is the degree of item overlap between items in the repeated and filler sequences[40]. The current results appear to extend Cumming et al.'s (2006) findings to sequences of nonwords.

On a related note, the current weak Hebb Effects may be related to the degree of experimental sensitivity. It may be that the current experimental design is simply not

---

[40] Note that this data only became available after Experiment 5 had been conducted.

sensitive enough to detect genuine effects of Hebb repetition; that is, the current design may inflate the chances of making a Type II error. Indeed, if one considers the patterns of means, there is a greater increase in recall performance over trials for Hebb compared to filler sequences for both word and nonword sequences. For example, performance increases from Trials 1 to 4 for Hebb sequences was 28% and 47% for word and nonword sequences, respectively. In contrast, performance increases for filler sequences were only 8% and 20% for word and nonword sequences, respectively. With this in mind, it is proposed that the current weak Hebb Effects may restrict the detection of genuine effects of phonological similarity on the learning of repeated sequences.

Finally, it is interesting to note that a number of previous studies have reported evidence for Hebb Effects when using less stringent criteria than has been adopted in Experiments 4 and 5 (e.g. Gagnon et al., 2004; Sechler & Watkins, 1991; Turcotte et al., 2005). These studies report Hebb Effects on the basis of a main effect of list type, with some studies finding no evidence for a list type x trials interaction (e.g. Sechler & Watkins, 1991; Turcotte et al., 2005) and others failing to even report the outcome of this interaction (e.g. Gagnon et al., 2004; Turcotte et al., 2005). In Experiments 4 and 5, evidence for clear and reliable Hebb Effects was defined by a main effect of list type and a list type x trials interaction. On this basis, the weak Hebb Effects in the present experiment may not be inconsistent with the Hebb Effects reported in previous studies.

Novel Prediction

The novel prediction that lowering demand on item learning would improve sequence learning compared to Experiment 4 was not supported. Nonword sequence learning failed to show any improvement from Experiment 4; learning reached 3.0 nonwords (50%) by Trial 4 in comparison to 2.8 nonwords (47%) in Experiment 4. Moreover, word sequences showed a considerable decrement in learning over trials compared with Experiment 4, reaching 4.0 words (57%) by Trial 4 compared with 5.3 (76%) in Experiment 4.

6.2.4 Summary

Experiment 5 generated a rather complex pattern of results. Firstly, lowering demand on item learning by presenting the same set of items for Hebb and filler lists failed to improve nonword sequence learning. A weak Hebb Effect was shown for word sequences, with even poorer evidence for a Hebb Effect for nonword sequences. It is tentatively hypothesised that these weak Hebb Effects may be due to the high degree of item overlap between Hebb and filler lists. Secondly, phonological similarity was shown to affect ISR performance for both word and nonword sequences, suggesting that participants utilise phonological coding during immediate recall of sequences. Thirdly, phonological similarity failed to impair the learning of repeated word sequences, suggesting that participants do not rely on PSTM to learn familiar material. The effect of phonological similarity on nonword sequence learning could not be reliably determined given the poor evidence for a Hebb Effect. It is therefore not clear whether participants utilised PSTM to learn sequences of unfamiliar material. Markov model analyses of between-trial learning and forgetting rates were of limited value given the poor sequence learning observed for both word and nonword repeated sequences. However, phonological similarity was shown to affect forgetting rates for both types of sequences, as in Experiment 4. This may provide further evidence, albeit somewhat limited, that sequences of similar items are more unstable and are susceptible to forgetting from trial to trial. Finally, as in Experiment 4, simple error analyses suggest that phonological similarity impairs the retention of order information for both word and nonword sequences.

In conclusion, the current results replicated a number of findings from Experiment 4. However, the effect of phonological similarity on nonword sequence learning is not clear in the present experiment. Experiment 6 therefore attempts a further replication of Experiment 4 and also addresses the limitations of Experiment 5 concerning the absence of reliable Hebb Effects and poor sequence learning, particularly for nonword sequences.

6.3 Experiment 6: Effects of Phonological Similarity and Hebb Repetition on Sequence Learning when Reducing Sequence Length

The primary aim of Experiment 6 was to attempt a further replication of Experiment 4. The results of Experiment 4 converged with the findings from paired-associate studies (e.g. Papagno & Vallar, 1992; Experiments 2 and 3) in providing support for the claim that PSTM mediates the long-term learning of unfamiliar material. Experiment 4 was therefore taken as preliminary evidence that the Hebb repetition paradigm may represent an analogue of the long-term learning of new phonological word-forms.

In light of Experiment 5's results, an important aim of Experiment 6 was to promote cumulative learning of repeated sequences. Given that reliable Hebb Effects were obtained in Experiment 4, the design of Experiment 6 reverted back to that adopted in Experiment 4, in which a different set of items was presented for the Hebb list and each of the two filler lists. Experiment 6 also incorporated two sets of materials in order to promote generalisability of the results. However, in an attempt to reduce problems associated with confirming parallel patterns of results across two sets of materials, Hebb and filler sequences were rotated across participants. Finally, given that the design used in Experiment 4 led to poor nonword sequence learning, the current experiment attempted to address this limitation by reducing nonword sequence length from six to five nonwords[41]. Word sequence length remained at seven words, given that word sequence learning was robust in Experiment 4.

The main predictions remained the same as in Experiments 4 and 5. One novel prediction was made. This predicted that reducing nonword sequence length would improve nonword sequence learning compared with Experiments 4 and 5.

6.3.1 Method

6.3.1.1 Design

Participants undertook an ISR task as described in Experiment 4 with one critical exception: nonword sequence length was reduced to five nonwords. The experiment used a 2 x 2 x 8 x 2 x 2 mixed-design incorporating three within-subjects factors: phonological similarity (similar items, distinct items), list type (Hebb, filler) and trials (1 to 8); and two between-subjects factors: lexicality (words, nonwords) and materials

---

[41] The decision to adopt a nonword sequence length of five was based on the results of a pilot experiment (n=6) in which nonword sequences contained four nonwords. Performance at Trial 1 reached an average of 2.9 nonwords (73%) for repeated sequences. It was assumed that this high level of performance at Trial 1 would restrict the opportunity for learning over trials to occur and would probably lead to ceiling effects. Indeed, no evidence for significant learning over trials was observed in this pilot. It was therefore decided to increase sequence length to five nonwords.

(materials A, materials B). As in Experiments 4 and 5, four experimental conditions were created: similar word sequences; distinct word sequences; similar nonword sequences; and distinct nonword sequences.

Participants were randomly assigned to one of the lexicality conditions. Within each lexicality condition, the factor of phonological similarity was counterbalanced across participants, as in Experiments 4 and 5. The two sets of materials devised in Experiment 4 were used in the current experiment. Participants were randomly divided so that half received materials A, and half materials B. For each set of materials within each experimental condition, the Hebb list (H) and each of the two filler lists (Filler 1, F_1; Filler 2, F_2) were rotated and served as F_1, F_2 and H lists for different participants. The assignment of each experimental set to F_1, F_2 and H was counterbalanced creating six different list orders for each experimental condition within each materials set. The presentation order of the list type factor was F_1, followed by F_2, followed by H. For each set of materials within each of the experimental conditions, the order of items within a list remained the same across participants.

6.3.1.2 Participants

Forty-eight undergraduate and postgraduate students attending the University of York volunteered to participate for course credit or payment. There were 37 females and 11 males aged between 18 years and 36 years (mean age of 19.9 years). All participants spoke English as their native language and reported no known hearing or language impairments.

6.3.1.3 Apparatus

The same apparatus was used as in Experiments 4 and 5.

6.3.1.4 Materials

The materials were the same as those used in Experiments 4 and 5. Each word set remained identical to that devised in Experiment 4.

Nonword Sets

One nonword from each of the 12 nonword sets devised in Experiment 4 was excluded, creating 5-item sets. Each of the 12 sets were matched on neighbourhood size using the CELEX database (Baayen et al., 1993; range across sets: 11.6 to 13.4). A one-way between-subjects ANOVA revealed no statistically significant differences between any of the 12 nonword sets (*F*(11, 48) = 0.08, *MSe* = 18.042, *p*=1.00). See Appendix 6(a) for the nonword sets.

Phonologically Similar and Distinct Nonword Sets: In each phonologically similar set, two different C_1 and three or four different C_2 were used. Each nonword within a set used the same vowel (*/e/* for materials A, */o/* for materials B). Each of the phonologically distinct sets used five different vowels. Each nonword within a set used a different C_1 and a different C_2.

Word Sets vs. Nonword Sets

Mean (SD) neighbourhood sizes values were 27.7 (5.9) and 12.4 (3.9) for word and nonword sets, respectively. A two-tailed independent t-test revealed a statistically significant difference between the two types of set (*t*(140.905) = 18.74, *p*<.0001, *r*=.84), confirming smaller neighbourhood size values for the nonword sets.

6.3.1.5 Procedure

The procedure and scoring criteria were identical to that described in Experiment 4. Maximum scores for each list were 7 for word sequences and 5 for nonword sequences.

6.3.2 Results

6.3.2.1 Correct Recall Analysis

Responses were scored following the same strict serial recall criterion as used in Experiment 1. No instances occurred in which a participant consistently made the same mispronunciation of an item. Instances in which the correct pronunciation of an item could not be clearly determined were scored as correct, provided they were recalled in the correct serial position; this occurred on eight occasions[42].

Data from one participant in the word conditions proved to be an outlier; this participant's data was excluded from all further statistical analyses[43]. The remaining data were normally distributed.

The means and standard deviations were firstly calculated for F_1, F_2, and H based on the 12 participants receiving materials A and the 11 participants receiving materials B for the word conditions, and the 12 participants receiving each of materials A and B for the nonword conditions. As in Experiments 4 and 5, mean scores were averaged over pairs of trials. In line with Experiment 4, to avoid conducting a complex five-way ANOVA for the main statistical analysis, these data were firstly subjected to separate three-way mixed-design and repeated-measures ANOVAs in order to confirm parallel patterns of results across (i) materials A and B for each of F_1, F_2, and H for each lexicality condition; and (ii) F_1 and F_2 for each lexicality condition. Each of these analyses revealed parallel patterns of results. The means and standard deviations were therefore calculated based on the 23 participants in the word conditions and the 24 participants in the nonword conditions, when collapsing across materials A and B, and across F_1 and F_2.

In keeping with the analyses performed in Experiments 4 and 5, the word and nonword data were analysed separately. Participants' scores were entered into two separate 2 x 2 x 4 repeated-measures ANOVAs each incorporating three within-subjects factors: phonological similarity (similar items, distinct items), list type (Hebb, filler), and learning trials (1 to 4).

The ANOVA conducted on the word data reported significant main effects of phonological similarity ($F(1,22) = 36.24$, $MSe = 2.931$, $p<.0001$, $r=.79$) and list type

---

[42] The word '*jam*' was perceived as '*gem*'; the word '*nut*' was perceived as '*knot*'; the word '*hum*' was perceived as '*ham*'; and the word '*col*' was perceived as '*coal*'. The nonword '*hom*' was perceived as '*hum*'; the nonword '*zep*' was perceived as '*sep*'; the nonword '*jus*' was perceived as '*jos*'; and the nonword '*gup*' was perceived as '*gop*'. These instances were presumably a consequence of accent differences between participants and the experimenter, although it is feasible that some, or all, of these instances may have been consistent mispronunciations.

[43] Outliers were defined as in Experiment 1 and calculated as described in Experiment 4.

($F(1,22) = 14.22$, $MSe = 6.805$, $p=.001$, $r=.63$), demonstrating better recall performance for distinct word sequences, and for Hebb lists. The main effect of trials also attained significance ($F(1.966,43.246) = 27.65$, $MSe = 0.947$, $p<.0001$, $\eta_p^2=.56$), confirming learning over trials (see Figure 6.5).

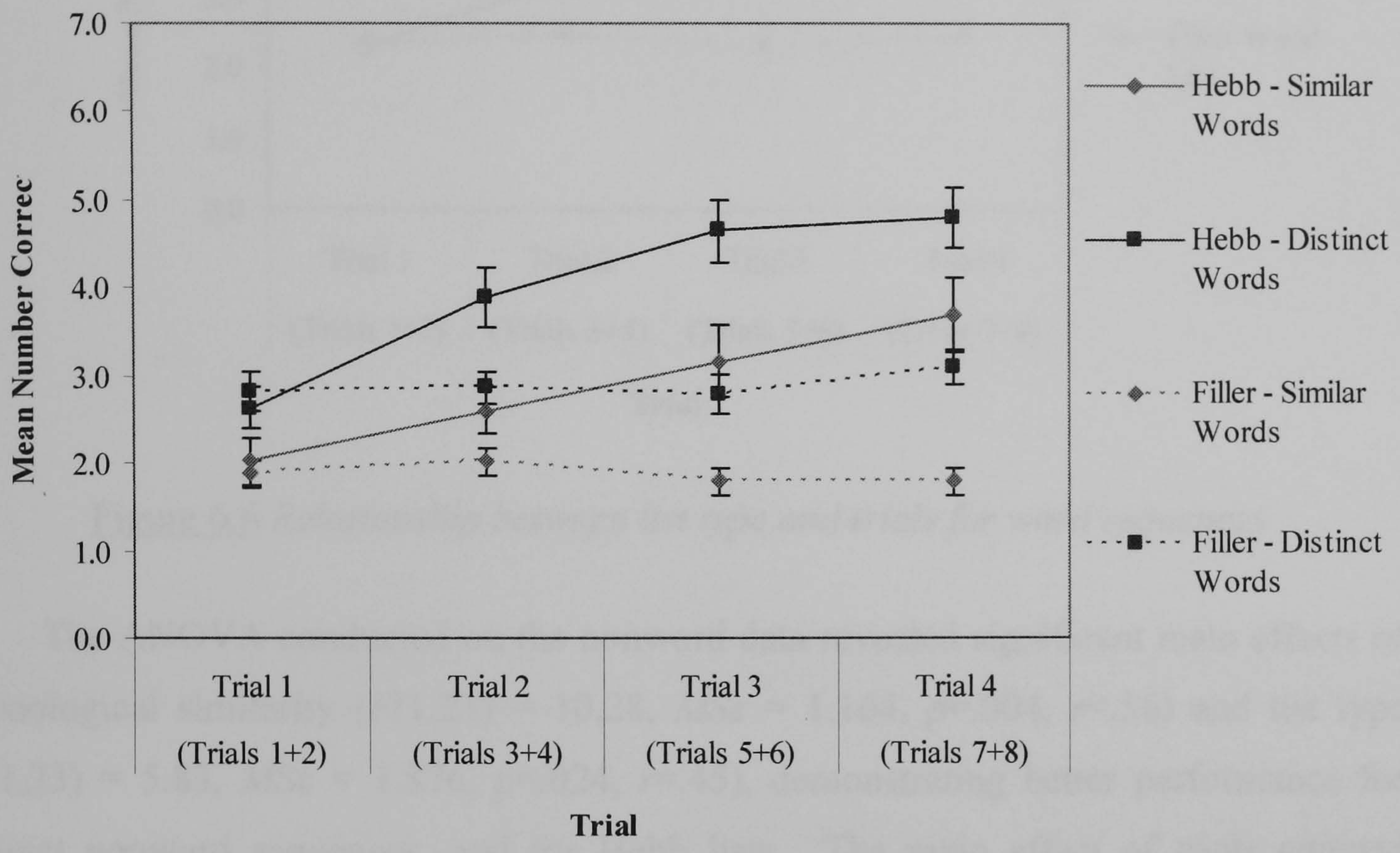


Figure 6.5: *ISR performance for word conditions as a function of phonological similarity, list type and trials (error bars represent standard error of the mean)*

A significant interaction between list type and trials emerged from the analysis ($F(3,66) = 47.51$, $MSe = 0.341$, $p<.0001$, $\eta_p^2=.68$), suggesting a differential pattern of learning over trials for Hebb and filler lists (see Figure 6.6). A simple main effects analysis yielded significant learning over trials for Hebb lists ($F(3,66) = 49.48$, $MSe = 0.67$, $p<.0001$) but not for filler lists ($F(3,66) = 0.96$, $MSe = 0.29$, $p=.42$). The remaining two-way interactions, and the three-way interaction, all failed to reach significance (all *ps*>.05).

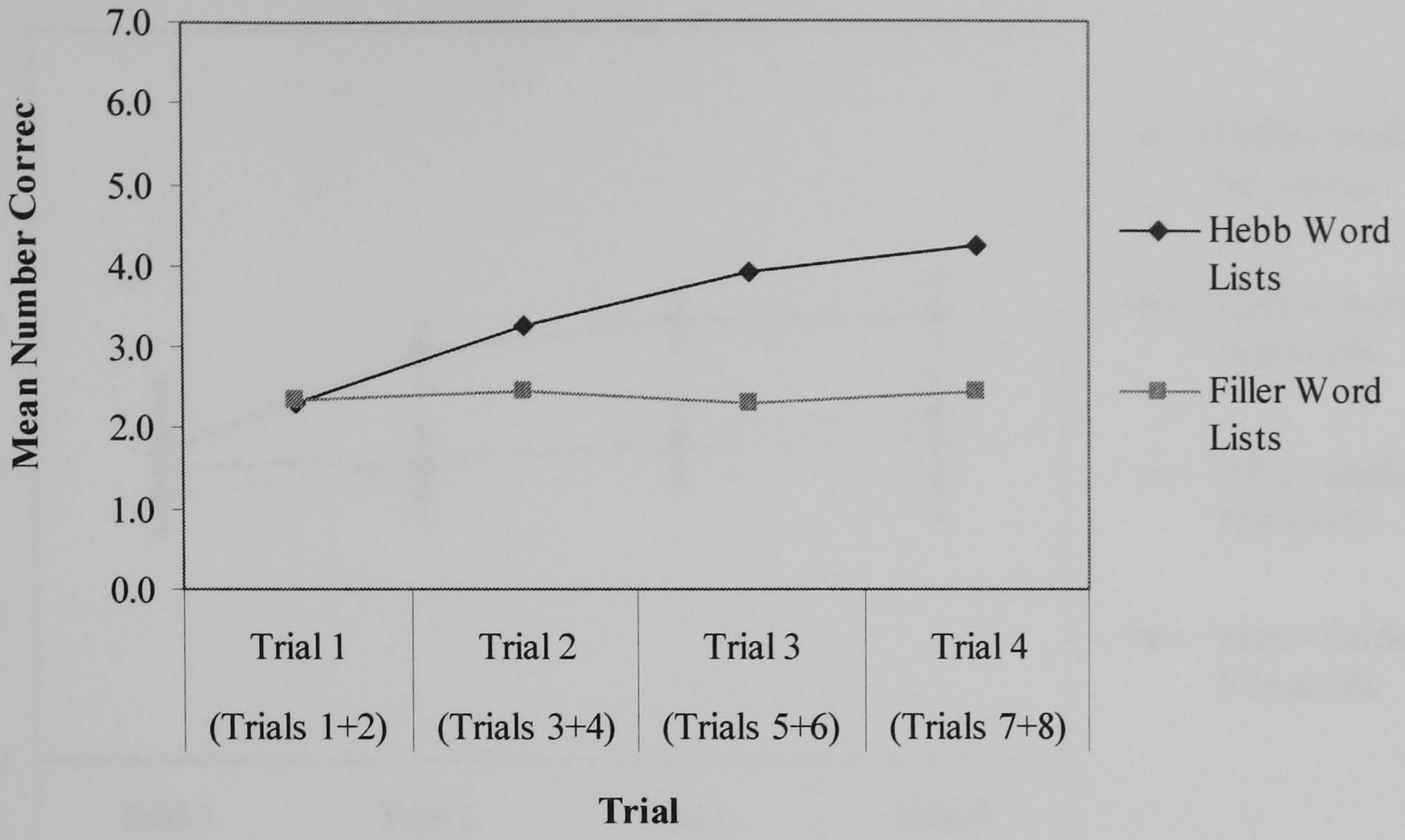


Figure 6.6 *Relationship between list type and trials for word sequences*

The ANOVA conducted on the nonword data revealed significant main effects of phonological similarity ($F(1,23) = 10.28$, $MSe = 1.164$, p=.004, r=.56) and list type ($F(1,23) = 5.83$, $MSe = 1.876$, p=.024, r=.45), demonstrating better performance for distinct nonword sequences, and for Hebb lists. The main effect of trials attained significance ($F(3,69) = 5.18$, $MSe = 0.374$, p=.003, η_p^2=.18), indicating learning over trials (see Figure 6.7).

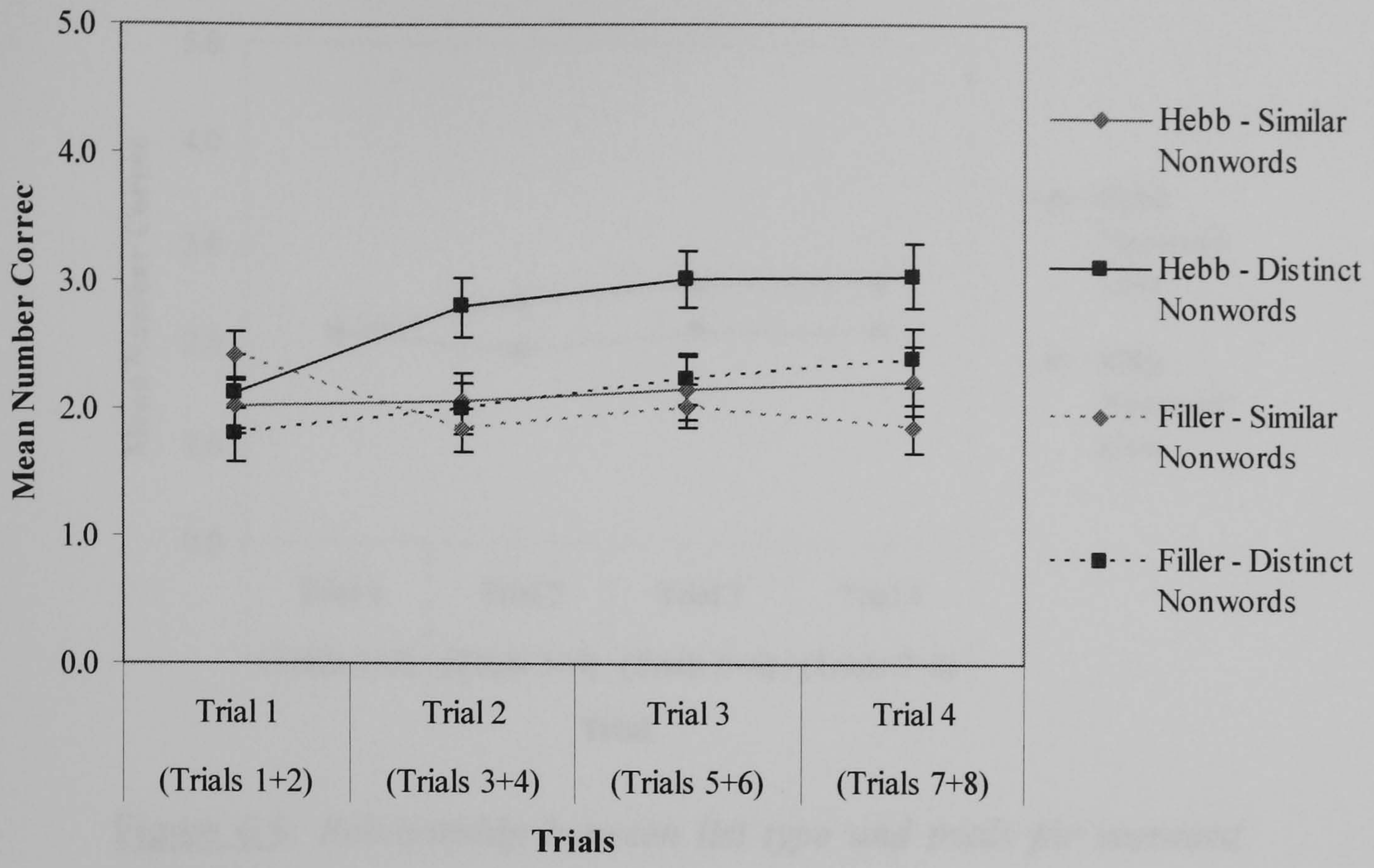


Figure 6.7: *ISR performance for nonword conditions as a function of phonological similarity, list type and trials (error bars represent standard error of the mean)*

The list type x trials interaction reached significance (*F*(3,69) = 3.52, *MSe* = 0.482, *p*=.019, η_p^2=.13), suggesting differential patterns of learning over trials for Hebb and filler lists (see Figure 6.8). A simple main effects analysis yielded significant learning over trials for Hebb lists (*F*(3,69) = 6.09, *MSe* = 0.52, *p*=.001) but not for filler lists (*F*(3,69) = 1.44, *MSe* = 0.34, *p*=.24).

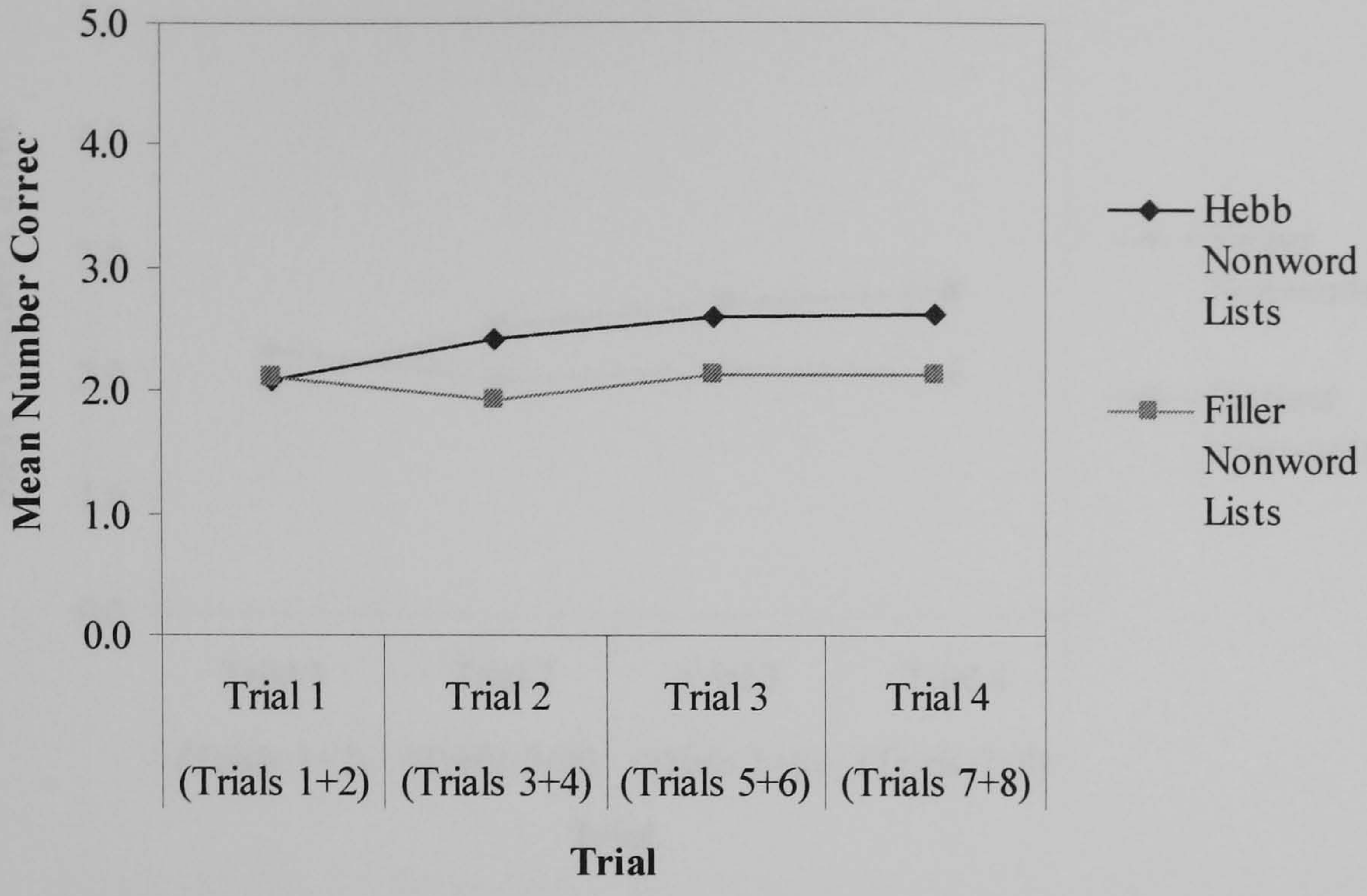


Figure 6.8: *Relationship between list type and trials for nonword sequences*

A significant phonological similarity x trials interaction emerged from the analysis (*F*(2.179,50.114) = 7.84, *MSe* = 0.723, *p*=.001, η_p^2=.25), demonstrating differential patterns of learning over trials for similar and distinct nonword sequences (see Figure 6.9). A simple main effects analysis yielded significant learning over trials for distinct nonword sequences (*F*(3,69) = 13.25, *MSe* = 0.41, *p*<.0001) but not for similar nonword sequences (*F*(3,69) = 1.23, *MSe* = 0.49, *p*=.30).

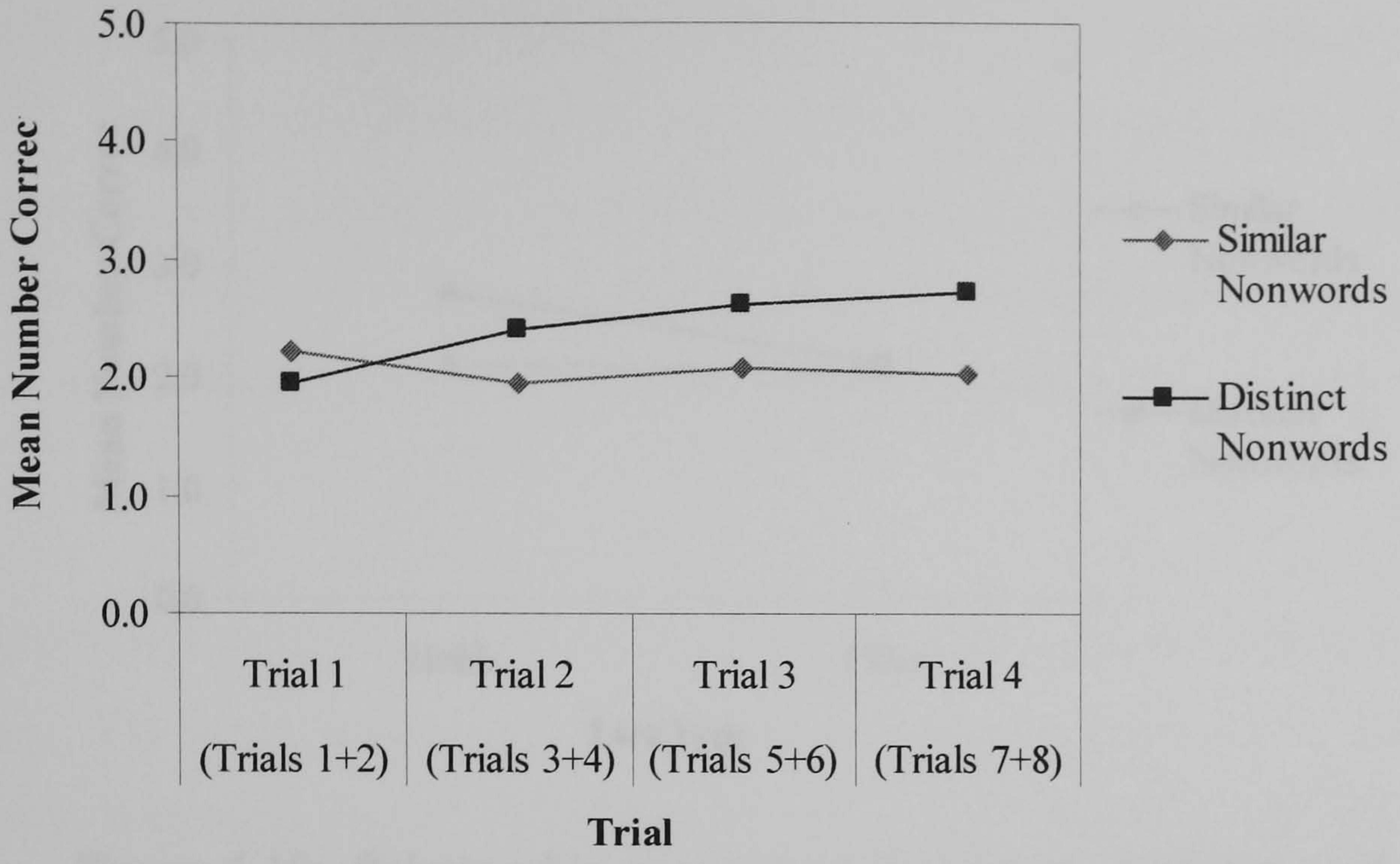


Figure 6.9: *Relationship between phonological similarity and trials for nonword sequences*

A marginally non-significant phonological similarity x list type interaction emerged from the analysis (*F*(1,23) = 3.79, *MSe* = 1.872, *p*=.064, *r*=.38). Given that this interaction attained significance in Experiments 4 and 5, it was subjected to a simple main effects analysis (see Figure 6.10). This analysis yielded a significant PSE for Hebb lists (*F*(1,23) = 8.79, *MSe* = 2.13, *p*=.007, *r*=.53) but not for filler lists (*F*(1,23) = 0.35, *MSe* = 0.90, *p*=.56). The three-way interaction failed to reach significance (*F*(3,69) = 0.65, *MSe* = 0.341, *p*=.58).

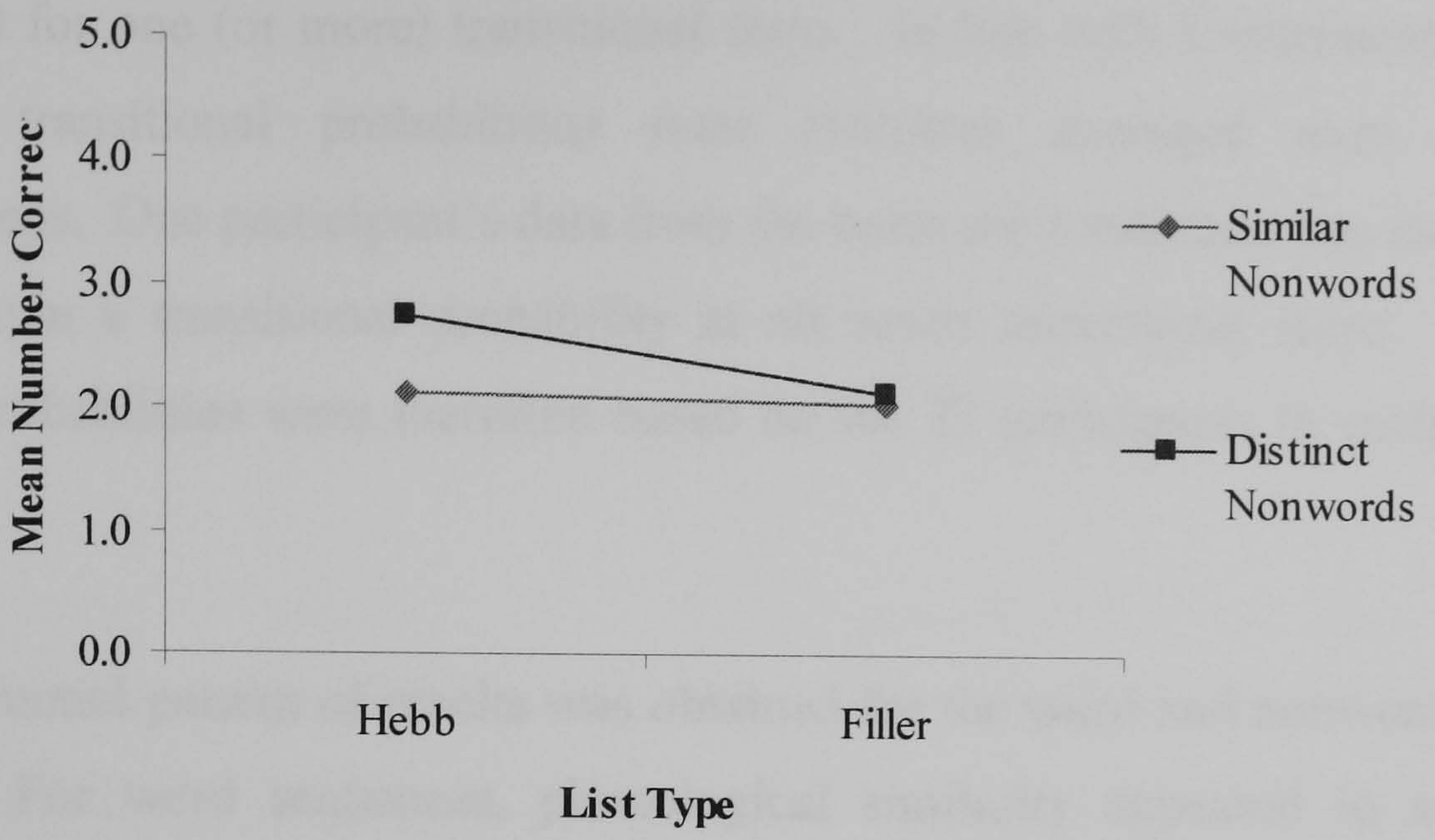


Figure 6.10: *Relationship between phonological similarity and list type for nonword sequences*

In summary, the results for word sequences showed clear and reliable evidence for a Hebb Effect and a PSE on ISR performance. Equivalent rates of learning were shown for similar and distinct word sequences. For nonword sequences, the results revealed clear and reliable evidence for a Hebb Effect and a PSE on ISR performance. Distinct repeated nonword sequences showed evidence of faster learning over trials compared to similar sequences.

6.3.2.2 Markov Model Analysis

To form a comparison with Experiments 4 and 5, analyses of learning and forgetting rates were conducted using an application of a Markov model in order to further investigate the degree of stability of items within a repeated sequence.

Learning and forgetting rates were calculated for the 23 participants in the word conditions and the 24 participants in the nonword conditions, following the same procedure as described in Experiment 4. Learning rates represent transitional probability (b) and forgetting rates represent transitional probability (d).

For each participant, transitional probabilities were calculated for each of the seven transitional steps following the same criteria as adopted in Experiments 4 and 5. As a result, 61% (n=14) and 88% (n=21) of participants in the word and nonword conditions, respectively, were excluded for failing to obtain transitional probabilities (b) and/or (d) for one (or more) transitional steps. In line with Experiments 4 and 5, participants' transitional probabilities were therefore averaged over all seven transitional steps. One participant's data from the nonword conditions was excluded for failing to obtain a transitional probability at all seven transitional steps. Averaged transitional probabilities were therefore based on the 23 participants in each lexicality condition.

A differential pattern of results was obtained for the word and nonword data (see Table 6.4). For word sequences, phonological similarity appeared to affect both learning and forgetting rates. Whereas distinct word sequences showed a higher learning rate, similar word sequences showed a higher forgetting rate. For nonword sequences, phonological similarity appeared to affect forgetting rates only. A higher forgetting rate was found for similar than distinct nonword sequences. Comparable learning rates were observed for similar and distinct nonword sequences.

Table 6.4: *Mean (and standard deviation) transitional probabilities for learning and forgetting transition types for similar and distinct word and nonword sequences*

	Condition			
	Word Sequences		Nonword Sequences	
Transition Type	Similar	Distinct	Similar	Distinct
Learning	.253 (.159)	.383 (.224)	.250 (.191)	.282 (.151)
Forgetting	.326 (.255)	.177 (.156)	.331 (.277)	.187 (.157)

The word and nonword data were analysed separately in line with the analyses based on correct recall performance. Participants' averaged transitional probability values were entered into 2 x 2 repeated-measures ANOVAs each incorporating two within-subjects factors: transition type (learning, forgetting) and phonological similarity (similar, distinct).

The ANOVA conducted on the word data reported non-significant main effects of transition type ($F(1,22) = 1.23$, $MSe = 0.082$, p=.28) and phonological similarity ($F(1,22) = 0.06$, $MSe = 0.035$, p=.81). The transition type x phonological similarity interaction attained significance (*F*(1,22) = 15.45, *MSe* = 0.029, *p*=.001, *r*=.64), suggesting differential learning and forgetting rates for similar and distinct word sequences. A simple main effects analysis yielded a significantly higher learning rate for distinct compared with similar word sequences (*F*(1,22) = 8.17, *MSe* = 0.02, *p*=.009, *r*=.52). In contrast, similar word sequences showed a significantly higher forgetting rate compared to distinct word sequences (*F*(1,22) = 6.39, *MSe* = 0.04, *p*=.019, *r*=.47).

For the nonword data, the ANOVA yielded non-significant main effects of phonological similarity (*F*(1,22) = 2.60, *MSe* = 0.028, *p*=.12) and transition type (*F*(1,22) = 0.02, *MSe* = 0.050, *p*=.88). A significant transition type x phonological similarity interaction emerged from the analysis (*F*(1,22) = 5.51, *MSe* = 0.032, *p*=.028, *r*=.45), suggesting differential patterns of learning and forgetting rates for the two types of sequence. A simple main effects analysis revealed a significantly higher forgetting rate for similar than distinct nonword sequences (*F*(1,22) = 6.15, *MSe* = 0.04, *p*=.021, *r*=.47) but equivalent learning rates for similar and distinct nonword sequences (*F*(1,22) = 0.53, *MSe* = 0.02, *p*=.47).

Further analyses were conducted on the word and nonword data in which missing transitional probabilities were replaced with transitional steps means. This was conducted in order to confirm that the data sets were free from bias due to the spread of missing data across transitional steps. These revealed an identical pattern of results to the initial analyses, thereby confirming the absence of any bias in the distribution of missing values.

In summary, the results for word sequences revealed a higher learning rate for distinct sequences and a higher forgetting rate for similar sequences. For nonword sequences, the results showed equivalent learning rates for similar and distinct sequences and a higher forgetting rate for similar sequences.

6.3.2.3 Error Analyses

In line with Experiments 4 and 5, detailed error analyses were conducted on the word and nonword data in order to investigate the degree of stability of items within a sequence. These analyses produced a complex pattern of results and failed to provide a coherent explanation of the interactive effects of phonological similarity and Hebb repetition on word and nonword sequence learning. As a result, the majority of these analyses are not reported here (but see Appendix 6(b) for tables showing the patterns of errors). However, to form a comparison with Experiments 4 and 5, simple error analyses were conducted on the two main error types (total item errors and order errors) in order to determine the overall effect of phonological similarity and Hebb repetition on word and nonword sequence learning

Participants' responses were scored and calculated for each of the two main error types following the same criteria as described in Experiment 4. Mean proportional values were collapsed across F_1 and F_2 and were averaged over every two trials. Order errors were calculated following the revised procedure: this lead to the exclusion of one participant's data from the word conditions and eight participants' data from the nonword conditions. Analyses were conducted on word and nonword data separately, as in Experiments 4 and 5.

Word Sequences: The pattern of errors indicates that phonological similarity and Hebb repetition affected the propensity of errors (see Table 6.5 and Appendix 6(b)).

Table 6.5: *Mean proportions (and standard deviations) for the two main error types for Hebb and filler lists collapsed over trials for similar and distinct word sequences*

	List Type/Error Type			
	Filler		Hebb	
Condition	Item	Order+	Item	Order+
Similar Words	.523 (.084)	.417 (.144)	.419 (.185)	.303 (.190)
Distinct Words	.484 (.103)	.200 (.117)	.364 (.175)	.096 (.102)

$^+$ Based on 22 participants

Participants' error scores, expressed as proportions, were entered into two separate 2 x 2 repeated-measures ANOVAs incorporating two within-subjects factors: phonological similarity (similar items, distinct items) and list type (Hebb lists. filler lists).

For total item errors, the ANOVA revealed significant main effects of phonological similarity (*F*(1,22) = 4.59, *MSe* = 0.011, *p*=.043, *r*=.42) and list type (*F*(1,22) = 9.97, *MSe* = 0.029, *p*=.005, *r*=.56), demonstrating a higher proportion of total item errors for similar sequences and for filler lists. The phonological similarity x list type interaction failed to reach significance (*F*(1,22) = 0.19, *MSe* = 0.008, *p*=.67).

For order errors, the ANOVA yielded significant main effects of phonological similarity (*F*(1,21) = 52.20, *MSe* = 0.019, *p*<.0001, *r*=.84) and list type (*F*(1,21) = 18.21, *MSe* = 0.014, *p*<.0001, *r*=.68), indicating a higher proportion of order errors for similar sequences and for filler lists. The phonological similarity x list type interaction failed to reach significance (*F*(1,21) = 0.07, *MSe* = 0.009, *p*=.79). An identical pattern of results were obtained when conducting the analysis in which missing order error proportions were replaced with trial means, thus confirming the absence of any bias in the spread of these missing values.

Nonword Sequences: The pattern of errors shows that phonological similarity and Hebb repetition affected the production of errors (see Table 6.6 and Appendix 6(b)).

Table 6.6: *Mean proportions (and standard deviations) for the two main error types for Hebb and filler lists collapsed over trials for similar and distinct nonword sequences*

	List Type/Error Type			
	Filler		Hebb	
Condition	Item	Order+	Item	Order+
Similar Nonwords	.420 (.117)	.260 (.122)	.409 (.193)	.210 (.158)
Distinct Nonwords	.553 (.180)	.081 (.106)	.433 (.205)	.046 (.089)

+ Based on 16 participants.

Participants' error scores, expressed as proportions, were entered into two separate 2 x 2 repeated-measures ANOVAs incorporating the same within-subjects factors as reported for the word sequences.

For total item errors, the ANOVA revealed significant main effects of phonological similarity (*F*(1,23) = 13.77, *MSe* = 0.011, *p*=.001, *r*=.61) and list type

($F(1,23) = 5.04$, $MSe = 0.020$, $p=.035$, $r=.42$), indicating a higher proportion of item errors for distinct sequence and for filler lists. The phonological similarity x list type interaction attained marginal significance ($F(1,23) = 4.26$, $MSe = 0.016$, $p=.050$. $r=.40$), suggesting a different pattern of phonological similarity on total item errors for Hebb and filler lists. A simple main effects analysis yielded a significantly higher proportion of total item errors for distinct than similar sequences for filler lists ($F(1,23) = 25.58$. $MSe = 0.01$, $p<.0001$, $r=.73$), but comparable proportions of these errors for similar and distinct sequences for Hebb lists ($F(1,23) = 0.36$, $MSe = 0.02$, $p=.55$).

For order errors, the ANOVA revealed a significant main effect of phonological similarity ($F(1,15) = 60.15$, $MSe = 0.008$, $p<.0001$, $r=.89$), demonstrating a higher proportion of order errors for similar sequences. The main effect of list type failed to reach significance ($F(1,15) = 2.20$, $MSe = 0.013$, $p=.16$) as did the phonological similarity x list type interaction ($F(1,15) = 0.22$, $MSe = 0.005$, $p=.65$). A second analysis in which missing order error proportions were replaced with trial means revealed an identical pattern of results to the initial analysis, thus confirming the absence of bias due to the spread of these missing values.

6.3.3 Discussion

Experiment 6 had two main aims. The first of these was to provide a replication of Experiment 4, in which evidence was generated in support of the PSTM hypothesis. The second aim was to address the limitations observed in Experiments 4 and 5 regarding the poor nonword sequence learning observed in both experiments and the absence of reliable Hebb Effects in Experiment 5. Experiment 6 attempted to address these limitations by reducing nonword sequence length to five nonwords and presenting a different set of items for the Hebb list and each of the filler lists, as in Experiment 4.

The main predictions were the same as in Experiments 4 and 5. One further prediction was made. This was that reducing nonword sequence length would improve nonword sequence learning compared to Experiments 4 and 5. The results of each prediction will be discussed for word and nonword sequences in turn.

Words

The analysis of correct recall performance supported the original prediction that a Hebb Effect would emerge for word sequences. Better overall performance was shown for Hebb compared to filler lists. Moreover, the list type x trials interaction confirmed cumulative learning over trials for repeated sequences, but not for filler sequences. This finding confirms that participants can benefit from the repeated presentation of an ordered sequence of words and thereby replicates Experiment 4 and previous studies reporting Hebb Effects for word sequences (e.g. Cumming et al., 2006; Gagnon et al., 2004; Page et al., 2006; Sechler & Watkins, 1991).

The original predictions that phonological similarity would impair ISR performance but not the learning of word sequences were upheld. Firstly, a PSE was observed on overall ISR performance. Furthermore, the size of the PSE was comparable for Hebb and filler lists, as indicated by the lack of a phonological similarity x list type interaction. Secondly, equivalent rates of learning similar and distinct word sequences were confirmed by the lack of both a significant phonological similarity x trials interaction and a significant three-way interaction. Indeed, parallel rates of learning were further supported by equivalent increases in performance over Trials 1 to 4 for similar (85% increase) and distinct (85% increase) repeated sequences. Moreover, these performance increases are comparable with those observed in Experiment 4 (74% and 89% increases for similar and distinct sequences, respectively).

These findings replicate those reported in Experiment 4, further confirming that participants utilise phonological coding during immediate recall of sequences but not when learning repeated sequences. The current results are also in line with recent studies (e.g. Fallon et al., 2005; Hitch et al., in press; Page et al., 2006) and the predictions of the Burgess and Hitch (1999) model of the phonological loop. Importantly, the present findings converge with the results of paired-associate studies (e.g. Papagno & Vallar, 1992).

In line with Experiments 4 and 5, a Markov model analysis of between-trial learning and forgetting rates was conducted on repeated word sequences. This produced an identical pattern of results to that reported in Experiment 4. Phonological similarity was shown to affect both between-trial learning and forgetting rates. Distinct sequences showed a higher learning rate coupled with a lower forgetting rate compared with similar sequences. This provides further support for the idea that the order of distinct words within a sequence is better maintained between trials compared to similar words; that is, distinct nonwords learned in particular positions on one trial are more likely to retain these positions on the next trial. In contrast, learning the order of a sequence of similar words appears to be more difficult, with similar words learned in particular positions on one trial being forgotten on the next trial. This suggests that similar word sequences are more unstable.

As was also found in Experiment 4, the Markov model analysis of between-trial learning and forgetting rates appears to contradict the results based on correct recall performance. Whereas better learning and less forgetting were observed for distinct sequences in the Markov model analysis, the correct recall performance analysis revealed equivalent learning over trials for these two types of sequence. However, this was shown in Experiment 4 to be due to differences in the amount of words to be learned between trials, with a smaller number of distinct words to be learned between trials leading to a higher transitional probability compared to similar words (see Tables 5.4 to 5.6, and Figure 5.6, on pp. 205-206 for a hypothetical example of this explanation). In sum, it is therefore concluded that the Markov model and correct recall performance analyses are not necessarily contradictory, but rather reflect differences in the way the data are examined. Moreover, the current Markov model analyses further confirms that similar word sequences are particularly fragile, with similar words being more likely to be forgotten between trials compared to distinct words.

As was found to be the case in Experiments 4 and 5, detailed error analyses revealed a complex pattern of results which were difficult to interpret. However, simple error analyses suggest that phonological similarity affected the overall production of order errors; a higher proportion of order errors were made for similar than distinct word sequences. This confirms the corresponding findings from Experiments 4 and 5 that phonological similarity impairs the immediate recall of a sequence of similar words, thus suggesting that recalling a sequence of similar words in the correct order is particularly difficult. The current result also supports earlier studies reporting a negative effect of phonological similarity on the retention of order information (e.g. Conrad, 1965; Henson et al., 1996; Wickelgren, 1965b). Phonological similarity was also shown to affect the overall production of total item errors as more of these errors were made for similar than distinct sequences. This suggests that participants are more likely to generate words not contained within a just-presented sequence of similar words. Finally, as in Experiment 4, Hebb repetition was shown to affect the production of both total item errors and order errors; a higher proportion of both types of errors were made for filler than Hebb lists. This may reflect the current experiment's robust Hebb Effect found for word sequences based on correct recall performance.

Nonwords

The analysis of correct recall performance provided support for the original prediction that a Hebb Effect would emerge for nonword sequences. Better overall ISR performance was shown for Hebb than filler lists. Moreover, cumulative learning over trials was observed for the repeated sequences only, as evidenced by a list type x trials interaction. This strengthens the results of Experiment 4 and confirms that participants can benefit from the repeated presentation of an ordered sequence of nonwords. Moreover, these results replicate previous studies which have reported Hebb Effects for nonword sequences (e.g. Gagnon et al., 2004; Turcotte, et al., 2005).

The original prediction that phonological similarity would impair ISR performance was confirmed. A PSE was observed on overall ISR performance. However, there was a trend to suggest that this PSE was confined to repeated sequences, as was found to be the case in Experiment 4. This pattern of results suggests that participants make use of phonological coding when recalling repeated nonword sequences. However, that a PSE was not observed on overall ISR performance for filler sequences may suggest that phonological similarity was not effective given that, on average, only two nonwords were recalled in the correct serial position over trials for similar and distinct filler sequences. Indeed, the finding that a PSE emerged on overall ISR performance for filler nonword sequences in Experiment 5 when a larger number of nonwords were recalled in the correct serial position over trials further supports this idea.

Phonological similarity was also shown to impair the learning of nonword sequences. The significant phonological similarity x trials interaction demonstrated significant learning over trials for distinct sequences but not for similar sequences.

Moreover, distinct Hebb sequences showed a larger increase in performance from Trials 1 to 4 compared with similar Hebb sequences (43% vs. 10% increase for distinct and similar Hebb sequences, respectively). However, the absence of a significant three-way interaction suggests that cumulative learning occurred over trials for distinct filler sequences as well as for distinct repeated sequences. Indeed, this idea is further supported by a 33% increase in performance over Trials 1 to 4 for distinct filler sequences. As such, this finding suggests that cumulative item learning can occur in the absence of repeated order information. A similar finding was also obtained in Experiment 5. The finding that phonological similarity disrupted the learning of repeated nonword sequences provides confirmatory evidence that, given the absence of existing lexical-semantic representations of nonwords in LTM, the learning of unfamiliar material relies upon phonological coding and as such is mediated by PSTM. This finding is consistent with the results of Experiment 4 and converges with the results of paired-associate studies (e.g. Papagno & Vallar, 1992) and Experiments 2 and 3.

A further interesting finding refers to the observation that the learning of distinct repeated nonword sequences shows an artificial asymptote; that is, learning appears to reach a 'glass ceiling' by Trial 3. This suggests that the majority of the nonword sequence learning observed takes place during the first four presentations of a repeated sequence[44]. Interestingly, that nonword sequence learning appears to occur predominately over the first three to four repeated presentations fits the findings from a previous study investigating Hebb Effects for nonwords (e.g. Turcotte et al., 2005). That learning does not appear to progress beyond this point may suggest that either participants are simply not able to learn all the nonwords within a sequence and/or that they are unable to learn the correct order of the nonwords within a sequence. Indeed, it may be that participants consistently produce the same errors over repeated presentations, despite the fact that the correct nonword sequence is repeatedly presented. This perseveration of errors may arise at the level of phonemes within nonwords and/or at the level of the order of whole nonwords within a sequence.

---

[44] Recall that performance is averaged over every two trials; therefore, Trials 3 and 4 represent performance for the fifth to eighth repeated presentations.

The analysis of between-trial learning and forgetting rates applying a Markov model analysis showed a higher forgetting rate for similar sequences, with comparable learning rates for similar and distinct sequences. This replicates the pattern of Markov model results obtained for nonword sequences in Experiment 4 and nonword pairs in Experiment 2. Thus, the current results can be seen to provide further evidence that learning the order of a sequence of similar nonwords between trials is particularly difficult and further confirms that sequences of similar nonwords are fragile. In contrast, once a distinct nonword has been learned in a particular position, it appears more likely to remain learned on subsequent trials. Furthermore, in line with the conclusions drawn in Experiments 2 and 4, the results of the Markov model analysis inform the results based on correct recall performance by highlighting that distinct nonwords are more likely to be retained between trials compared to similar nonwords.

As was proposed in Experiment 4, the fragility of similar nonword sequences may also represent a degree of fragility at the level of phonemes. It may be that the phonemes contained with a similar nonword are harder to bind together to form a whole nonword; as such, the higher between-trial forgetting rate shown for similar nonword sequences may indicate that similar nonwords are more likely to fall apart between trials compared with distinct nonwords.

Finally, although detailed error analyses failed to shed further insights regarding the effects of phonological similarity and Hebb repetition on nonword sequence learning, simple error analyses demonstrated that phonological similarity impairs the recall of an ordered sequence of similar nonwords. In line with Experiments 4 and 5, this suggests that recalling a sequence of similar nonwords is more difficult than recalling a sequence of distinct nonwords. This finding also converges with previous studies (e.g. Gathercole et al., 2001). Phonological similarity also affected the production of total item errors. As was found in Experiment 4, a higher proportion of total item errors were produced for distinct than similar nonword sequences. However, in contrast to Experiment 4, this was found to be the case for filler lists only; comparable proportions of total item errors were made for similar and distinct sequences for Hebb lists. This suggests that participants more often recall nonwords from outside a just-presented filler sequence when that sequence contains distinct nonwords rather than similar nonwords. Finally, Hebb repetition affected the generation of total item errors, as shown by a higher proportion of these errors for filler compared to Hebb lists. In contrast, Hebb repetition did not affect the production of order errors. Given that Experiments 4 and 5 failed to find an effect of Hebb repetition on both total item errors and order errors, the current finding may reflect the reliable Hebb Effect based on correct recall performance found in the current experiment.

Hebb Effect Issues

Of further interest is the finding that stronger evidence was found for Hebb Effects for word and nonword sequences in the present experiment and Experiment 4, in comparison to the weak Hebb Effects for both types of sequence obtained in Experiment 5. This latter finding has been tentatively attributed to the degree of item overlap between items in the Hebb and filler sequences, as was also argued by Cumming et al. (2006). The finding that reliable Hebb Effects emerged in Experiments 4 and 6 provides further support for this idea, given that both experiments adopted a design whereby there was no item overlap between Hebb and filler sequences. Moreover, this idea is further strengthened by the broadly convergent pattern of Hebb Effects shown for word and nonword sequences.

Results for Novel Prediction

The novel prediction that reducing nonword sequence length would improve nonword sequence learning was not supported. Nonword sequence learning reached 3.0 nonwords (60%) by Trial 4, representing only a 10-13% increase from Experiments 4 and 5. It is therefore concluded that nonword sequence learning remains relatively poor in the current experiment despite the reduction in nonword sequence length.

6.3.4 Summary

Experiment 6 generated a pattern of results broadly consistent with Experiment 4. Firstly, clear and reliable Hebb Effects were found for word and nonword sequences, confirming that participants are able to benefit from the repeated presentation of an ordered sequence of items. Secondly, phonological similarity was shown to affect ISR performance for word and nonword sequences, demonstrating that participants utilise phonological coding during immediate recall of sequences of familiar and unfamiliar material. However, the lack of a PSE on overall ISR performance for filler nonword sequences suggests that two few nonwords were recalled in order for an effect of phonological similarity to emerge. Thirdly, and importantly, phonological similarity selectively disrupted the learning of nonword sequences, but not the learning of word sequences. Furthermore, in contrast to Experiment 4, cumulative learning was observed for both distinct repeated and filler nonword sequences, suggesting that item learning can occur without repeated order information. Taken together, these findings provide converging evidence that PSTM is relied upon when learning unfamiliar material, but not when learning familiar material and, as such, provide some support for the PSTM hypothesis. Moreover, the current results are consistent with paired-associate studies (e.g. Papagno & Vallar, 1992) and Experiments 2 and 3. However, nonword sequence learning remained relatively poor despite reducing nonword sequence length. The Markov model analysis of between-trial learning and forgetting rates revealed an identical pattern of results to Experiment 4 for both word and nonword sequences. Moreover, the current Markov model results parallel those reported for nonword pair learning in Experiment 2. Phonological similarity affected between-trial forgetting rates for both word and nonword sequences, further confirming that sequences containing similar items are unstable between trials. Taken together, these findings suggest that phonological similarity influences the rate at which similar nonwords are forgotten between trials. Finally, simple error analysis revealed that phonological similarity increases the propensity of order errors for both word and nonword sequences, suggesting that phonological similarity impairs the retention of order information.

In conclusion, Experiment 6 provided a close replication of Experiment 4. To this end, it is proposed that Experiments 4 and 6 provide limited evidence to suggest that the Hebb repetition paradigm may possibly represent an analogue to the long-term learning of new word-forms. However, it is important to acknowledge that this latter claim is rather tentative at this stage on the basis that relatively poor learning of nonword sequences was observed in Experiments 4 and 6. Experiment 7 therefore offers a further attempt at improving nonword sequence learning in an attempt to provide more substantial support for this claim.

6.4 Experiment 7: Effects of Phonological Similarity and Hebb Repetition on Nonword Sequence Learning using Serial Order Reconstruction

The main aim of Experiment 7 was to improve nonword sequence learning. Nonword sequence performance in Experiments 4 and 6 reached between 2.8 nonwords (47%) and 3.0 nonwords (60%) by Trial 4, respectively. In contrast, word sequence performance by Trial 4 had reached between 4.8 words (69%) in Experiment 6 and 5.3 words (76%) in Experiment 4.

Experiment 5 failed to improve nonword sequence learning when presenting the same set of items for the Hebb and filler lists; moreover, this design may have eliminated the Hebb Effects originally observed in Experiment 4. Experiment 6 reverted back to presenting a different set of items for the Hebb and filler lists but nevertheless failed to improve nonword sequence learning, despite reducing nonword sequence length to five nonwords.

Given that reliable Hebb Effects appear to emerge only when there is no item overlap between the Hebb and filler lists, demand on nonword learning remains relatively high when presenting different sets of items for Hebb and filler lists. An alternative method of improving nonword sequence learning then might be to present a task which does not require the recall of item information; that is, a method which primarily requires the retention of order information only. The serial recognition (SR) and serial order reconstruction (SOR) tasks both rely to a great extent on memory for order information (e.g. Gathercole et al., 2001; Nimmo & Roodenrys, 2005; Saint-Aubin & Poirier, 1999). In the SR task, two sequences of items are presented, separated by a brief interval. On some trials, these two sequences are identical; on other trials, the order of two adjacent items is transposed in the second sequence. The participant's task is to indicate whether the items in the two sequences were presented in the same serial order. In the SOR task, participants are presented with a sequence of items; at recall, these items are re-presented and the participant's task is to reconstruct the order in which the items were originally presented.

The demand on item learning is therefore considerably lower in the SR and SOR tasks compared to the ISR task. The ISR task requires the storage of both item and order information, given that the items have to be recalled from memory in the absence of these items at test. In contrast, the SR and SOR tasks primarily rely on the ability to retain order information due to the re-presentation of the items at test.

Experiment 7 therefore maintained the design of Experiments 4 and 6, by presenting a different set of items for the Hebb list and each of the filler lists in order to facilitate a Hebb Effect, but adopted the recall method of SOR. In addition, nonword sequence length was set at six nonwords, as in Experiment 4, based on the expectation that participants may show better levels of performance at the start of learning than in previous experiments given the low demand on item learning. It is important to consider the potential effects of Hebb repetition and phonological similarity on recall performance and sequence learning when using this alternative task. Firstly, given that participants have been shown to be capable of benefiting from the repeated presentation of an ordered sequence of nonwords in Experiments 4 and 6, Hebb Effects should occur using the SOR method given that this task requires the retention of order information. Indeed, Page et al. (2006) found evidence of Hebb Effects for sequences of letters, nameable pictures and words using the method of SOR.

Secondly, phonological similarity should still impair the recall of individual sequences on the basis that this variable is known to impair the retention of order information (Gathercole et al., 2001), as evidenced by an increase in the frequency of order errors (e.g. Conrad, 1965; Henson et al., 1996; Wickelgren, 1965b). Indeed, a number of studies have reported effects of phonological similarity on recall performance using tasks which require the retention of order information only (e.g. Henson, Hartley, Burgess, Hitch & Flude, 2003; Gathercole et al., 2001; Nimmo & Roodenrys, 2005; Thorn, Gathercole & Frankish, 2002). Gathercole et al. (2001) report strong and comparable degrees of phonological similarity for ISR and SR tasks. Moreover, phonological similarity has been shown to impair recall performance to a similar extent for words and nonwords using the SR task (Gathercole et al., 2001; Nimmo & Roodenrys, 2005).

A further aim of Experiment 7 then was to provide a replication of Experiments 4 and 6. These experiments reported Hebb Effects for nonword sequences. Moreover, phonological similarity was shown to selectively impair ISR performance and the learning of nonword sequences. This latter result suggests that participants utilise phonological coding, and thus rely on PSTM, to learn unfamiliar material. An effect of phonological similarity on the learning of nonword sequences may still be expected when using the SOR method of recall on the basis that participants still have to rely on temporary representations of the phonological structure of nonwords in order to create permanent representations in LTM.

To recap, a number of predictions were made. Firstly, it was hypothesised that adopting the alternative task of SOR would improve nonword sequence learning compared with Experiments 4 and 6. Secondly, a Hebb Effect was predicted. Thirdly, it was predicted that phonological similarity would impair nonword recall performance. Finally, phonological similarity was predicted to disrupt the learning of nonword sequences, as evidenced by slower learning of similar sequences.

6.4.1 Method

6.4.1.1 Design

Participants undertook a serial order reconstruction (SOR) task in which sequences of phonologically similar and distinct nonwords were auditorily presented. Participants were then required to immediately recall aloud the items in their correct serial order from a visual array of just-presented sequence items. The Hebb list and each of the two filler lists each comprised a different set of items. The experiment used a 2 x 2 x 8 within-subjects design incorporating three within-subjects factors: phonological similarity (similar nonwords, distinct nonwords), list type (Hebb, filler) and trials (1 to 8). Two experimental conditions were created: similar nonword sequences; distinct nonword sequences.

The factor of phonological similarity was counterbalanced as in previous experiments. For each experimental condition, three different sets of items were constructed: one for the Hebb list (H) and one for each of the two filler lists (Filler 1, F_1; Filler 2, F_2). These sets were rotated and served as F_1, F_2 and H lists for different participants. The assignment of each experimental set to F_1, F_2 and H was counterbalanced creating six different list orders for each experimental condition. The presentation order of the list type factor was F_1, followed by F_2, followed by H. For each set of materials within each of the experimental conditions, the order of items within a list and the order of items within each visual array remained the same across participants.

6.4.1.2 Participants

Twenty-four undergraduate students attending the University of York volunteered to participate for course credit or payment. There were 17 females and 7 males aged between 18 years and 29 years (mean age of 20.0 years). All participants spoke English as their native language and reported no known hearing or language impairments.

6.4.1.3 Apparatus

The same apparatus was used as in Experiments 4 to 6.

6.4.1.4 Materials

The materials were the nonwords used in Experiments 4 to 6.

Nonword Sets

Six sets of six nonwords were devised: three phonologically similar sets and three phonologically distinct sets. Each of the six sets were matched on neighbourhood size using the CELEX database (Baayen et al., 1993; mean range across sets: 12.3 to 13.2). A one-way between-subjects ANOVA revealed no statistically significant differences between the six nonword sets ($F(5, 30) = .03$, $MSe = 16.372$, $p=1.00$). See Appendix 7(a) for the nonword sets.

Phonologically Similar Nonword Sets: These were the sets used as materials A in Experiments 4 and 5[45]. Each nonword within a set had the same vowel (*/e/)*. Each set used two different C_1 and four different C_2

---

[45] Recall that 5-item nonword sets were used in Experiment 6.

Phonologically Distinct Nonword Sets: These were constructed from materials A and B in Experiments 4 to 6[46]. Each set used five different vowels. Each nonword within a set used a different C_1 and a different C_2.

Construction of Nonword Visual Arrays

For each experimental condition, eight visual arrays were created for each of F_1, F_2 and H. No two visual arrays were alike. Each visual array contained the six nonwords within a set, printed in size 48 font on sheets of white paper. The six nonwords formed a circle in the middle of the sheet (see Appendix 7(b) for an example of a visual array).

6.4.1.5 Procedure

The procedure closely followed that described in Experiment 4 but with one modification regarding the method of recall. Participants were advised that they would hear sequences containing six nonwords which they would be required to immediately recall aloud from a visual array of nonwords. They were advised that immediately preceding the presentation of each sequence, a sheet of paper containing the nonwords to-be-presented would be placed directly in front of them face down. Participants were advised that they were to turn the sheet over at the beginning of each recall period and use it as a prompt to recall the nonwords. Participants were instructed to say 'blank' if they could not remember the nonword that had appeared in a particular position.

The scoring criteria were the same as described in Experiment 4. The maximum score was 6 for each sequence.

---

[46] This was done so as to eliminate five of the nonwords used in Experiments 4 to 6 which were later identified as real, although highly infrequent, words (e.g. *hom, jus, fid, vas, gid, som*). The lexical status of these 'words' was tested in a lexical decision experiment: 18 participants were presented with a written version all the materials used in Experiments 4 to 6 (words and nonwords were randomly intermixed) with the task of indicating whether each item was a word or nonword. The task was not time-constrained. All of these real words were classified as nonwords by all participants. This confirmed that the nonword conditions in Experiments 4 to 6 were not contaminated by the presence of real words.

6.4.2 Results

6.4.2.1 Correct Recall Analysis

Responses were scored following the same strict serial recall criterion as described in Experiment 1. No instances occurred in which a participant consistently made the same mispronunciation of a nonword. Instances in which the correct pronunciation of a nonword could not be clearly determined were scored as correct, provided they were recalled in the correct serial position; this occurred on one occasion[47].

The mean scores obtained for each participant were calculated for each of F_1, F_2 and H for each of the two conditions. The data were normally distributed and no outliers were detected[48]. In line with Experiments 4 to 6, mean scores were averaged over every two trials. In order to confirm a parallel pattern of results across F_1 and F_2, the data were subjected to a repeated-measures ANOVA. This revealed an identical pattern of results for F_1 and F_2 (all *p*s>.05), allowing these data to be pooled for all further statistical analyses.

Participants' scores were entered into a three-way repeated-measures ANOVA incorporating three within-subjects factors: phonological similarity (similar, distinct), list type (Hebb, filler) and trials (1 to 4). The ANOVA revealed significant main effects of phonological similarity ($F(1,23) = 98.58$, $MSe = 2.354$, $p<.0001$, $r=.90$) and trials ($F(3,69) = 3.96$, $MSe = 0.727$, $p=.012$, $\eta_p^2=.15$), confirming better recall performance for distinct nonword sequences, and learning over trials (see Figure 6.11). The main effect of list type failed to attain significance ($F(1,23) = 0.62$, $MSe = 2.028$, $p=.44$).

---

[47] The nonword 'vom' was perceived as 'vum', presumably due to accent differences between the participant and the experimenter.

[48] Outliers were defined and calculated as detailed in Experiment 4.

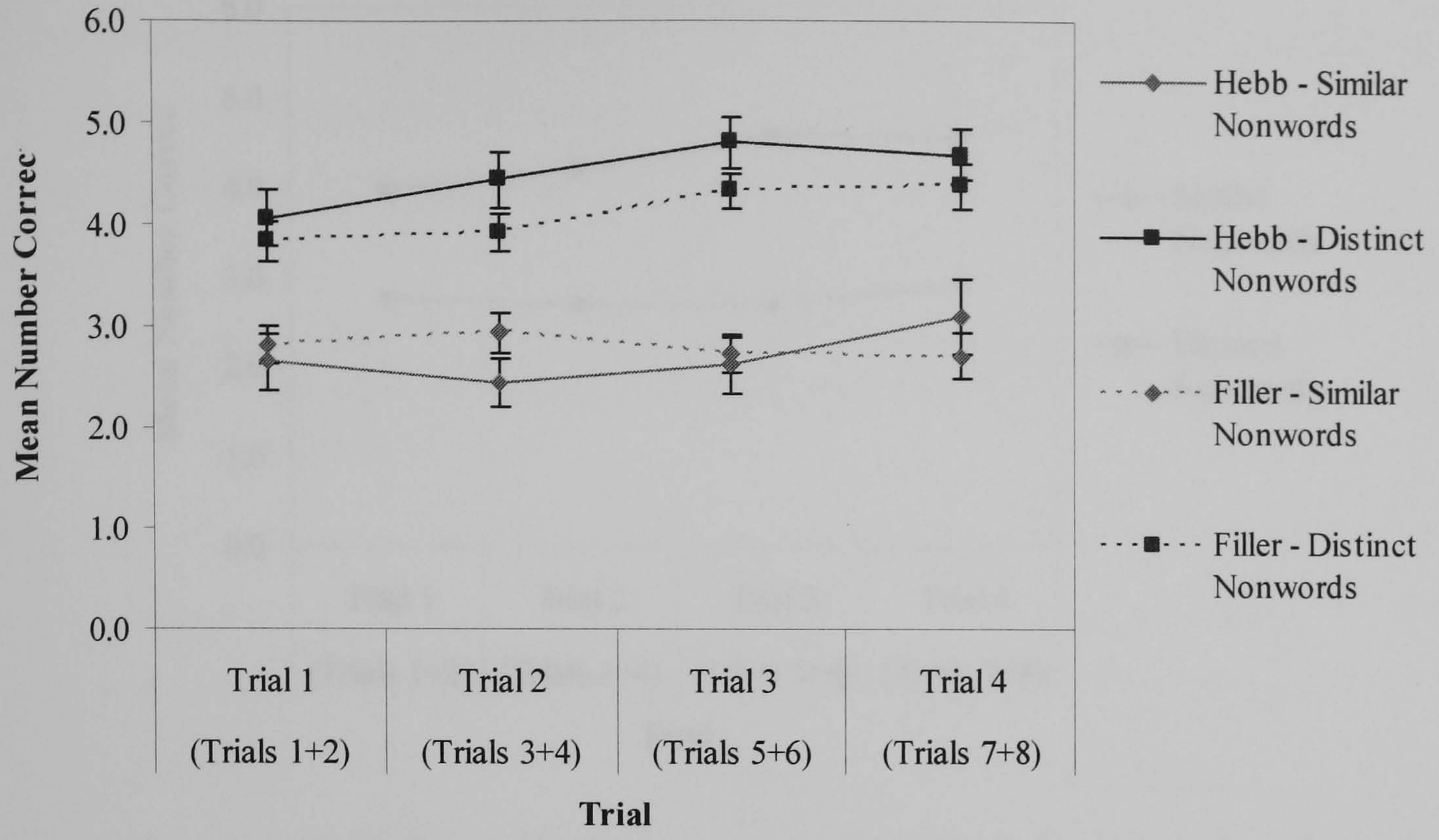


Figure 6.11: *Recall performance for nonword sequences as a function of phonological similarity, list type and trials (error bars represent standard error of the mean)*

The phonological similarity x trials interaction approached significance ($F(3,69) = 2.67$, $MSe = 0.706$, $p=.054$, $\eta_p^2=.10$), suggesting differential patterns of learning over trials for similar and distinct sequences. Given that this interaction attained significance in Experiments 4 to 6, it was subjected to a simple main effects analysis (see Figure 6.12). This yielded significant learning over trials for distinct sequences ($F(3,69) = 7.12$, $MSe = 0.59$, $p<.0001$) but not for similar sequences ($F(3,69) = 0.66$, $MSe = 0.84$, $p=.58$). The remaining two-way interactions, and the three-way interaction, all failed to reach significance (all *p*s>.05).

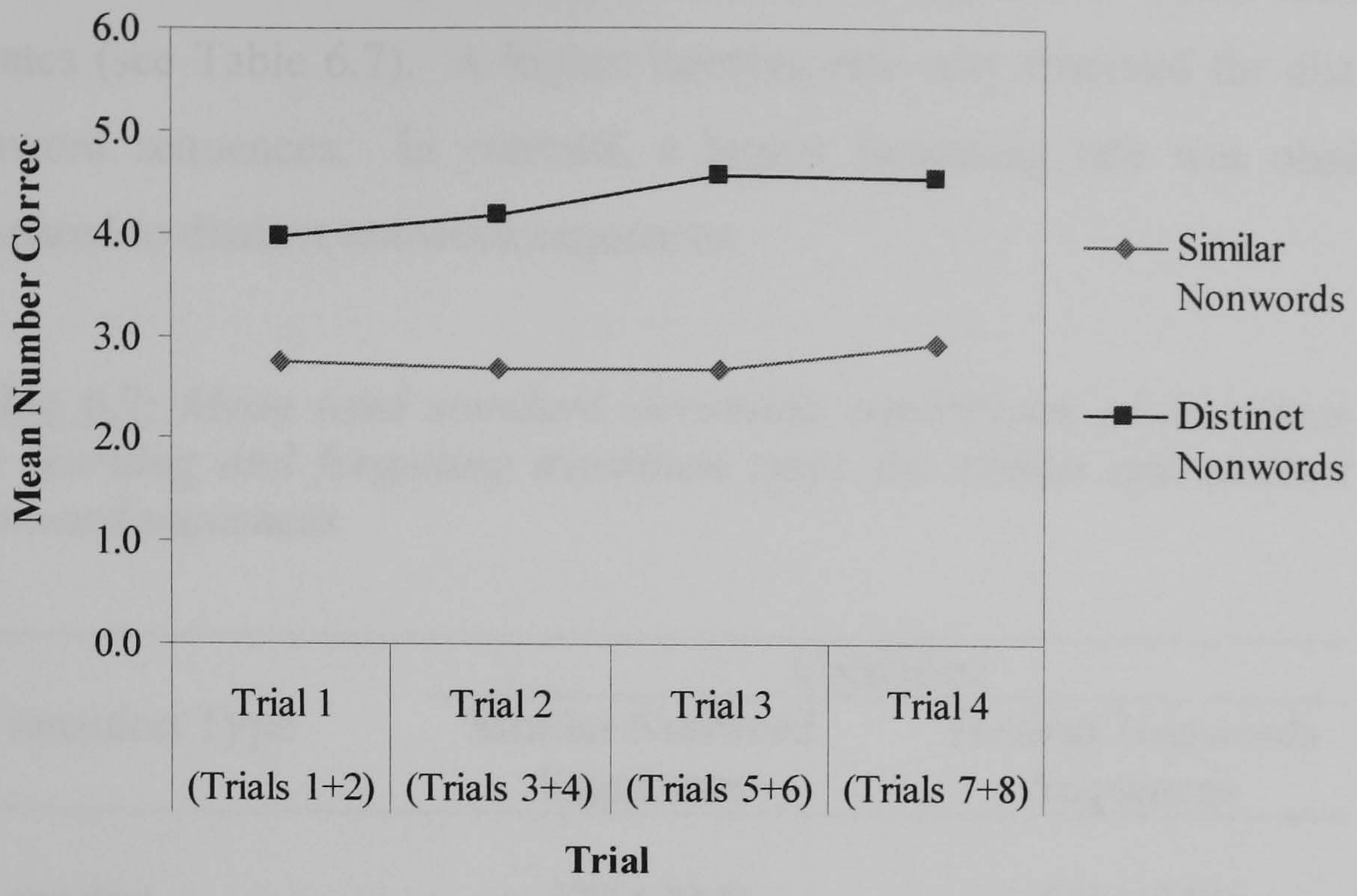


Figure 6.12: *Relationship between phonological similarity and trials*

6.4.2.2 Markov Model Analysis

Analyses of learning and forgetting rates were conducted using an application of a Markov model in order to investigate the degree of stability of items within a repeated sequence and to form a comparison with Experiments 4 to 6.

Learning and forgetting rates were calculated for the 24 participants following the same procedure as described in Experiment 4. Learning rates represent transitional probability (b) and forgetting rates represent transitional probability (d).

For each participant, transitional probabilities were calculated for each of the seven transitional steps as in Experiments 4 to 6. Following this criteria, 92% (n=22) of participants were excluded for failing to obtain transitional probabilities (b) and/or (d) for one (or more) transitional steps. In line with Experiments 4 to 6, participants' transitional probabilities were subsequently averaged over all seven transitional steps. As a result, one participant's data was excluded for failing to obtain a transitional probability at all seven transitional steps. Averaged transitional probabilities were therefore based on the 23 participants.

The results showed that phonological similarity appears to affect learning and forgetting rates (see Table 6.7). A higher learning rate was observed for distinct than similar nonword sequences. In contrast, a higher forgetting rate was observed for similar compared to distinct nonword sequences.

Table 6.7: *Mean (and standard deviation) transitional probabilities for learning and forgetting transition types for similar and distinct nonword sequences*

Transition Type	Condition	
	Similar Nonword Sequences	Distinct Nonwords Sequences
Learning	.322 (.216)	.646 (.344)
Forgetting	.327 (.199)	.151 (.148)

Participants' averaged transitional probability values were entered into a two-way repeated-measures ANOVA incorporating two within-subjects factors: transition type (learning, forgetting) and phonological similarity (similar, distinct). The ANOVA reported a significant main effect of transition type ($F(1,22) = 14.22$, $MSe = 0.098$, $p=.001$, $r=.63$), showing higher overall learning rates. The main effect of phonological similarity failed to reach significance ($F(1,22) = 2.60$, $MSe = 0.047$, $p=.12$). The transition type x phonological similarity attained significance ($F(1,22) = 29.93$, $MSe = 0.048$, $p<.0001$, $r=.76$), indicating a differential pattern of learning and forgetting rates for similar and distinct sequences. A simple main effects analysis yielded a significantly higher learning rate for distinct nonword sequences ($F(1,22) = 16.62$, $MSe = 0.07$, $p=.001$, $r=.66$) but a significantly higher forgetting rate for similar nonword sequences ($F(1,22) = 15.35$, $MSe = 0.02$, $p=.001$, $r=.64$).

6.4.2.3 Error Analyses

To form a comparison with Experiments 4 to 6, simple error analyses were conducted in order to investigate the overall effect of phonological similarity and Hebb repetition on nonword sequence learning. However, in contrast to Experiments 4 to 6, these error analyses were only conducted on order errors given the low frequency of occurrence of total item errors (but see Appendix 7(c) for tables showing the patterns of errors).

Participants' responses were scored and calculated for order errors following the same revised procedure as described in Experiment 4. No participants were excluded following this revised procedure.

Table 6.8 shows that phonological similarity, and to a lesser extent Hebb repetition, had an affect on the production of order errors.

Table 6.8: *Mean proportions (and standard deviations) for the two main error types for Hebb and filler lists collapsed over trials for similar and distinct nonword sequences*

	List Type/Error Type			
	Filler		Hebb	
Condition	Item	Order	Item	Order
Similar Nonwords	.030 (.041)	.503 (.116)	.028 (.072)	.524 (.193)
Distinct Nonwords	.043 (.054)	.279 (.111)	.048 (.067)	.214 (.166)

Participants' error scores, expressed as proportions, were entered into a 2 x 2 repeated-measures ANOVAs incorporating two within-subjects factors: phonological similarity (similar items, distinct items) and list type (Hebb lists, filler lists).

The ANOVA revealed a significant main effect of phonological similarity (F(1,23) = 99.70, MSe = 0.017, p<.0001, r=.90), demonstrating a higher proportion of order errors for similar sequences. The main effect of list type failed to reach significance (F(1,23) = 0.79, MSe = 0.015, p=.38) as did the phonological similarity x list type interaction (F(1,23) = 3.87, MSe = 0.011, p=.06).

6.4.3 Discussion

Experiment 7 had two aims. The first of these was to improve the poor nonword sequence learning observed in Experiments 4 and 6. This was attempted by adopting a

SOR task. This task primarily requires the retention of order information and so reduces the demand on item learning. The second aim was to provide a replication of Experiments 4 and 6. These experiments found evidence for Hebb Effects for nonword sequences. Moreover, phonological similarity was shown to impair both ISR performance and the learning of repeated nonword sequences.

The predictions were firstly that reducing demand on item learning by adopting an SOR task would improve nonword sequence learning compared to Experiments 4 and 6. Secondly, it was predicted that a Hebb Effect would emerge. Finally, phonological similarity was predicted to impair nonword recall performance and the learning of nonword sequences.

The prediction that adopting a SOR task would improve nonword sequence learning was not supported. Although performance reached 4.7 nonwords (78%) by Trial 4, performance at Trial 1 was considerably higher compared to Experiments 4 to 6. This latter finding was anticipated given the low demand on item learning in the SOR task. As a result, nonword sequence learning over trials was restricted. Indeed, the increase in performance over Trials 1 to 4 was considerably lower (15% increase) compared to Experiment 4 (115% increase) and Experiment 6 (43% increase).

The analysis of correct recall performance provided very little support for the original prediction that a Hebb Effect would emerge for nonword sequences. Comparable levels of overall recall performance were found for Hebb and filler lists. Although the main effect of trials demonstrated that learning did occur over trials, the absence of a significant list type x trials interaction suggests that cumulative learning occurred over trials to a similar extent for Hebb and filler sequences. As such, this finding fails to support the Hebb Effects for nonword sequences observed in Experiments 4 and 6.

The absence of a reliable Hebb Effect in the present experiment was surprising. It was proposed in Experiment 5 that the Hebb Effect is influenced by the degree of item overlap between Hebb and filler lists. A reliable Hebb Effect was therefore expected in the present experiment given that there was no item overlap between the items in the Hebb and filler lists. Considering these points, it is proposed that the present poor Hebb

Effect may be a consequence of the change in recall method from ISR to SOR. It may be that the reduced requirement for item learning in the SOR task reduces the normally beneficial effect of repeated order information. This may suggest that participants need to explicitly learn the nonwords themselves for cumulative learning of a sequence of nonwords to occur. That is, participants may need to produce nonwords at recall on the basis of temporary representations within PSTM rather than relying on the knowledge that the nonwords will be re-presented at test. Indeed, the recent finding that a Hebb Effect occurs for sequences of familiar material (e.g. letters, nameable pictures and words) when adopting the method of SOR (e.g. Page et al., 2006) may suggest that an important part of learning a sequence of nonwords is acquiring the phonological forms of each nonword. The lack of a reliable Hebb Effect in the present experiment may also reflect the use of different strategies when adopting the SOR task compared to the ISR task. Given that the nonwords are re-presented at test in the current experiment, participants may find it easier to utilise strategies, such as remembering the nonwords on the basis of the first or final phoneme rather than the whole nonword, in order to recall a nonword sequence. The use of such a strategy may reduce the benefit of repeated order information due to reliance on the visual presentation of the nonwords at recall. Alternatively, it may also be that the process of comparing a phonological representation of a nonword constructed at presentation with the written form of that nonword presented at test interferes with PSTM for order information. Finally, it may be that the current experimental design is simply not sensitive enough to detect genuine effects of Hebb repetition. Whatever the explanation, the current results suggest that a task which relies primarily on the retention of order information, such as SOR, is not particularly conducive to learning a repeated sequence of nonwords.

The results generated support for the original prediction that phonological similarity would impair nonword recall performance. A PSE was observed on overall recall performance. Furthermore, the size of the PSE was shown to be comparable for Hebb and filler lists. This provides evidence that participants utilise phonological coding during the immediate recall of repeated and filler nonword sequences. However, that a PSE on recall performance was shown for filler nonword sequences contrasts with the findings from Experiments 4 and 6, which reported a PSE on ISR performance for repeated nonword sequences only. It is suggested that this difference across experiments is related to the number of items recalled in the correct serial position for filler nonword sequences. That is, phonological similarity may not be effective when too few nonwords are recalled in the correct serial position. With this in mind, a PSE on recall performance for filler nonword sequences would therefore be expected in the current experiment given that between three and four nonwords were recalled in the correct serial position, as opposed to only two nonwords in Experiments 4 and 6.

Finally, it was not possible to determine the genuine effect of phonological similarity on nonword sequence learning due to the limited evidence for a Hebb Effect. Although marginally better learning over trials was shown for distinct compared to similar sequences; the absence of a significant three-way interaction suggests that this pattern of learning was the same for repeated and filler nonword sequences. However, in contrast to these statistical results, Figure 6.11 (p. 262) suggests that equivalent rates of learning were shown for similar and distinct repeated sequences. Indeed, comparable increases in performance over Trials 1 to 4 were found for similar (19%) and distinct (15%) repeated sequences. Interesting, the current results show a similar pattern to the corresponding results from Experiment 5, in which a poor Hebb Effect was also obtained. To this end, it is concluded that the role of PSTM in the learning of unfamiliar material cannot be reliably established in the absence of reliable Hebb Effects.

The Markov model analysis of between-trial learning and forgetting rates conducted on the repeated nonword sequences demonstrated that phonological similarity affected both learning and forgetting rates. Similar nonword sequences showed a higher between-trial forgetting rate coupled with a lower between-trial learning rate. This suggests that similar nonword sequences are harder to learn and retain from trial to trial. However, this pattern differs somewhat from Experiments 4 and 6. Although similar nonword sequences also showed higher between-trial forgetting rates, comparable between-trial learning rates were obtained in Experiments 4 and 6. It is proposed that this difference in between-trial learning rates reflects the poor nonword sequence learning based on correct recall performance and the poor Hebb Effect observed in the current experiment. Interestingly, the current pattern of Markov model results is in line with the corresponding results from Experiment 5, in which it was also concluded that the Markov model analyses were restricted due to poor nonword sequence learning and a poor Hebb Effect.

Finally, simple error analyses revealed that phonological similarity impaired the recall of nonword sequences; a higher proportion of order errors were produced for similar than distinct sequences. In line with Experiments 4 to 6, this suggests that learning the order of a sequence of similar nonwords is particularly difficult and thus confirms that phonological similarity impairs the retention of order information (e.g. Gathercole et al., 2001). Hebb repetition failed to affect the production of order errors as indicated by comparable proportions of these errors for Hebb and filler lists. This finding may possibly reflect the current experiment's lack of a reliable Hebb Effect based on correct recall performance.

6.4.4 Summary

Experiment 7 generated an unexpected pattern of results. Firstly, the use of a SOR task failed to improve nonword sequence learning compared with Experiments 4 and 6. Evidence was found for only a very weak Hebb Effect, despite presenting a different set of items for Hebb and filler lists. As such, it is suggested that participants may need to explicitly learn the phonological forms of nonwords by producing them from PSTM in order to form stable representations of these in LTM. Phonological similarity was shown to impair nonword recall performance, confirming that participants utilise phonological coding during immediate recall of nonword sequences. The results regarding the effect of phonological similarity on the learning of nonword sequences were difficult to interpret due to the absence of a reliable Hebb Effect, as was found to be the case in Experiment 5. As such, it is not clear whether phonological similarity had a genuine effect on the learning of nonword sequences. In a similar vein, the results of the Markov model analysis were also restricted due to poor nonword sequence learning and the absence of a reliable Hebb Effect. However, simple error analyses confirmed that phonological similarity impairs the retention of order information for sequences of nonwords.

In conclusion, the results of Experiment 7 did not provide a complete replication of Experiments 4 and 6. As such, Experiment 7 therefore offers no firm support for the PSTM hypothesis.

6.5 General Discussion

The experiments presented in this chapter had an overarching aim: to further investigate the PSTM hypothesis by providing a converging pattern of results across paired-associate and Hebb repetition tasks. Initial evidence was provided in support of this aim in Experiment 4 (Chapter Five). Hebb Effects were observed for word and nonword sequences, demonstrating that the Hebb repetition paradigm can be used to investigate rates of learning. Importantly, phonological similarity was shown to selectively disrupt the learning of nonword sequences, but not the learning of word sequences. The results of Experiment 4 therefore provide limited evidence that the Hebb repetition paradigm might be an analogue to the long-term learning of novel phonological word-forms. Experiments 5, 6 and 7 therefore attempted to replicate the results of Experiment 4. A second aim of these experiments was to address the main limitation observed in Experiment 4; that of poor nonword sequence learning.

The patterns of results generated from Experiments 5, 6 and 7 were rather complex. However, a number of conclusions can be drawn by considering the patterns of word and nonword sequence results across the three experiments. Consider first the emergence of Hebb Effects. Reliable Hebb Effects were obtained for word and nonword sequences only when the design adopted an ISR task involving no item overlap between Hebb and filler lists (Experiment 6). This suggests that the Hebb Effect is influenced by the degree of item overlap between Hebb and filler lists. It seems that participants are more able to benefit from repeated order information when the interference from filler sequences is reduced by presenting a different set of items for Hebb and filler lists. Furthermore, the lack of a reliable Hebb Effect for nonword sequences in Experiment 7 may highlight the importance of producing nonwords from temporary phonological representations of nonwords in PSTM in promoting the cumulative learning of repeated nonword sequences.

A broadly consistent pattern of results was obtained regarding the effect of phonological similarity on ISR performance. Significant PSEs were observed on ISR performance for word sequences (Experiments 5 and 6) and nonword sequences (Experiments 5, 6 and 7). This demonstrates that participants utilise phonological coding during the immediate recall of sequences of both familiar and unfamiliar material and supports the findings of Experiment 4. However, there was some evidence in Experiment 6 to suggest that phonological similarity is not effective when too few items are recalled in the correct serial position, as was shown for filler nonword sequences. Indeed, this conclusion is supported by a similar finding in Experiment 4.

The pattern of word sequence learning over trials was very consistent across Experiments 5 and 6. In both of these experiments, phonological similarity failed to impair the learning of word sequences, evidenced by equivalent rates of learning similar and distinct sequences. This consistent pattern of findings replicates Experiment 4 and provides further support for the claim that PSTM does not mediate the learning of familiar material. Instead, participants are able to rely on existing lexical-semantic representations of words in LTM in order to learn word sequences. Importantly, these findings are consistent with the results of paired-associate studies (e.g. Papagno & Vallar, 1992) and to a limited degree Experiments 2 and 3.

In contrast to the consistent pattern of word sequence learning, nonword sequence learning was shown to be more varied over Experiments 5, 6 and 7. Phonological similarity was found to reliably disrupt nonword sequence learning only in Experiment 6. Faster learning over trials was observed for distinct repeated sequences compared to similar repeated sequences. This finding replicates Experiment 4. However, in contrast to Experiment 4, Experiment 6 provided additional evidence that cumulative learning also occurs for distinct filler sequences, suggesting that item learning can occur without the need for repeated order information. Taken together, Experiments 4 and 6 suggest that the learning of unfamiliar material is mediated by PSTM and thus provide converging evidence with paired-associate studies (e.g. Papagno & Vallar, 1992) and Experiments 2 and 3.

In contrast, Experiments 5 and 7 failed to provide conclusive evidence regarding the effect of phonological similarity on nonword sequence learning given that only very weak Hebb Effects were obtained in these experiments. It is therefore concluded that the absence of reliable Hebb Effects restricts the detection of any genuine effect of phonological similarity on nonword sequence learning. Although the nonword sequence learning results in Experiments 5 and 7 have highlighted the importance of obtaining robust Hebb Effects, these results are considered limited in terms of their value and informativeness for this thesis.

The analyses of between-trial learning and forgetting rates conducted on repeated sequences using an application of a Markov model produced a rather varied pattern of results over Experiments 5, 6 and 7. For word sequences, phonological similarity was shown to affect forgetting rates in both Experiments 5 and 6, suggesting that similar word sequences are more unstable and susceptible to forgetting between trials compared to distinct word sequences. However, the effect of phonological similarity on between-trial learning rates differed across Experiments 5 and 6. Whereas equivalent learning rates were observed for similar and distinct sequences in Experiment 5, a higher learning rate was shown for similar sequences in Experiment 6. The pattern of results shown in Experiment 6 suggests that distinct word sequences are better learned and retained between trials compared to similar word sequences; that is, distinct words are more likely to be retained from trial to trial. In contrast, similar word sequences appear to be more unstable, with different similar words being learned between trials. This pattern of results is consistent with Experiment 4. The Markov model analysis results in Experiment 5 were considered of limited value on the basis of poor word sequence learning over trials and the presence of a weak Hebb Effect, as shown by the correct recall performance results.

The pattern of between-trial learning and forgetting rates for nonword sequences was also relatively inconsistent across Experiments 5, 6 and 7. Phonological similarity was shown to affect between-trial forgetting rates across all three experiments, demonstrating that similar nonword sequences are particularly fragile between trials compared to distinct nonword sequences. However, the effect of phonological similarity on between-trial learning rates differed across the three experiments. Higher learning rates were found for distinct nonword sequences in Experiments 5 and 7, with comparable learning rates shown for similar and distinct sequences in Experiment 6. However, the Markov model results in Experiments 5 and 7 were restricted due to poor nonword sequence learning and poor Hebb Effects, based on correct recall performance. As such, the Markov model results in Experiment 6 were considered to present a more reliable account of between-trial learning and forgetting rates. The pattern of results shown in Experiment 6 suggests that learning a sequence of similar nonwords is particularly difficult, with similar nonwords learned in particular positions on one trial being forgotten on the next trial. This fragility for similar nonword sequences may occur at the level of the order of nonwords within a sequence and/or at the level of individual phonemes within nonwords. Finally, the pattern of results shown in Experiment 6 is consistent with Experiment 4 and is also in line with the corresponding results for nonword pair learning in Experiment 2

It is therefore concluded that analyses of between-trial learning and forgetting rates using a Markov model provide additional insights into the effect of phonological similarity on sequence learning, but only when reliable sequence learning is observed. In particular, these analyses have highlighted the role of between-trial forgetting in learning word and nonword sequences, suggesting that similar word and nonword sequences are particularly fragile and susceptible to forgetting between trials. Such information is not available when considering the results based on correct recall performance alone.

The results of simple error analyses revealed a relatively consistent pattern of results over Experiments 5 and 6. For word sequences, the findings from the analyses conducted on order errors showed that phonological similarity impairs the recall of similar word sequences in both Experiments 5 and 6. This suggests that recalling a sequence of similar words in the correct order is particularly difficult. This pattern of results is consistent with Experiment 4. Indeed, this finding may somewhat support the results of the Markov model analyses conducted on repeated word sequences in Experiments 4 and 6. Phonological similarity was also shown to increase the production of total item errors for similar word sequences in Experiments 5 and 6, suggesting that participants are more likely to recall a word from outside the pool of words in a just-presented sequence when that sequence contains similar words. Hebb repetition increased the production of both total item errors and order errors for filler compared to Hebb lists in Experiment 6. This pattern of results is consistent with Experiment 4. In contrast, Hebb repetition increased the production of only total item errors for filler lists in Experiment 5. The effect of Hebb repetition on the production of errors in Experiments 5 and 6 may be related to the extent to which a Hebb Effect emerges based on correct recall performance; indeed, a reliable Hebb Effect was observed for word sequences in Experiments 4 and 6, but not in Experiment 5.

The pattern of simple error analyses for nonword sequences revealed a somewhat varied pattern of results across Experiments 5, 6 and 7. Phonological similarity was shown to increase the production of order errors for similar sequences across all three experiments. This suggests that recalling the order of a sequence of nonwords is more difficult when that sequence contains similar compared to distinct nonwords. This pattern of results is somewhat reminiscent of the results of the Markov model analyses in Experiments 4 and 6. Indeed, this pattern of findings suggests that fragility for similar nonword sequences does occur at the level of the order of nonwords within a sequence. Turning to total item errors, phonological similarity differentially affected the production of this error type in Experiments 5 and 6[49]. Whereas a higher proportion of total item errors were made for similar filler nonword sequences in Experiment 5, a higher proportion of these errors were made for distinct filler nonword sequences in Experiment 6. However, comparable proportion of total item errors were shown for similar and distinct Hebb nonword sequences in both Experiments 5 and 6. Finally, Hebb repetition failed to affect the production of both total item errors and order errors in Experiment 5 and order errors in Experiment 7. This may reflect the lack of reliable Hebb Effects for nonword sequences based on correct recall performance in Experiments 5 and 7. In contrast, Hebb repetition increased the production of total item errors for filler lists but not the production of order errors in Experiment 6.

Finally, Experiments 5, 6 and 7 attempted to improve the poor nonword sequence learning originally observed in Experiment 4 by reducing demand on item learning in a number of different ways. Experiment 5 reduced the number of unique items to learn across sequences, Experiment 6 reduced nonword sequence length and Experiment 7 adopted a SOR task. However, none of these design modifications succeeded in improving nonword sequence learning. Possible reasons for this consistent pattern of very limited nonword sequence learning will be discussed in Chapter Seven.

---

[49] Recall that total item errors were not analysed in Experiment 7 due to their low frequency of occurrence.

Chapter Seven: General Discussion and Conclusions

7.1 Overview

The overarching aim of this thesis was to extend our current understanding of the role of PSTM in the long-term learning of novel phonological word-forms. This aim was investigated by conducting further tests of the PSTM hypothesis using two learning paradigms: paired-associate and Hebb repetition tasks. A further aim was to examine whether the Hebb repetition paradigm can be viewed as an experimental analogue to the long-term learning of new phonological word-forms.

This chapter will begin by providing a brief review of the research that inspired the present experiments. It will then present a summary of the main empirical findings, before discussing their theoretical interpretations and the extent to which they provide support for the PSTM hypothesis. The broader implications of the findings will also be considered. The chapter will then move on to examine the extent to which the Hebb repetition paradigm can be seen as an analogue of new word-form learning. A number of methodological issues arising from the research will then be highlighted. Limitations of the current research and some ideas for future work will be discussed, before offering some overall conclusions.

7.2 Background Research

The research conducted in this thesis was based on the claim that PSTM plays a crucial role in the process by which novel phonological word-forms become stable and permanent representations within LTM (e.g. Baddeley et al., 1998). Chapter One reviewed a considerable amount of research which provided converging evidence in support of the PSTM hypothesis (e.g. Baddeley, 1993; Baddeley et al., 1988; Cheung, 1996; Gathercole & Baddeley, 1989a, 1990a, 1990b; Gathercole et al., 1992; 1997, 1999; Masoura & Gathercole, 1999, 2005; Papagno et al., 1991; Papagno & Vallar, 1992; Service 1992; Service & Kohonen, 1995; Thorn & Gathercole, 1999; 2001).

Previous research has been primarily governed by developmental and experimental word learning studies conducted with children (e.g. Cheung, 1996; Gathercole & Baddeley, 1989a, 1990a, 1990b; Gathercole et al., 1992; 1997, 1999; Masoura & Gathercole, 1999, 2005; Michas & Henry, 1994; Service 1992; Service & Kohonen, 1995). This is not particularly surprising given that childhood represents an intensive period of vocabulary acquisition. A further way of investigating the complex processes and mechanisms involved in vocabulary acquisition has been to observe the effects on learning of manipulating variables known to interfere with the operation of the phonological loop in clearly defined ways; such variables include word length, phonological similarity and articulatory suppression (e.g. Baddeley et al., 1975; Conrad & Hull, 1964; Murray, 1967, 1968). As discussed in Chapter One, the concept of the phonological loop has proved capable of accounting for each of these PSTM effects (e.g. Baddeley, 1986). With this in mind, studies which manipulate PSTM variables may provide a direct test of the PSTM hypothesis. That is, if PSTM plays a crucial role in the long-term learning of unfamiliar material, then variables that are known to affect the operation of the phonological loop should reveal corresponding effects on the learning of this material.

However, on reviewing the literature, it became apparent that only two studies have conducted such a test of the PSTM hypothesis (e.g. Papagno et al., 1991; Papagno & Vallar, 1992). Both studies were conducted with adults and investigated the effects of articulatory suppression (e.g. Papagno et al., 1991), word length and phonological similarity (e.g. Papagno & Vallar, 1992) on the paired-associate learning of familiar and unfamiliar material. In each study, the manipulated variable disrupted the learning of word-nonword pairs, but not the learning of word-word pairs. These findings were taken as evidence in support of the PSTM hypothesis. However, although such studies may offer important insights into the role of PSTM in new word-form learning if they were to be conducted with children, it is not always appropriate to manipulate PSTM variables when this type of task is given to children. Presumably, children may find it difficult to learn more than two or three word-nonword pairs (e.g. Gathercole et al., 1997). As a result, the manipulation of PSTM variables such as word length and phonological similarity would not be appropriately implemented given this small number of word-nonword pairs. To this end, it was argued that experimental word learning studies conducted with adults may offer additional insights into the detailed underpinnings of the process of vocabulary acquisition by permitting a more intensive and detailed investigation of new word-form learning abilities than can be conducted with children. In turn, this allows for a richer collection of data which can then be analysed at a more in-depth level. As such, the first aim of this thesis was to provide a further test of the PSTM hypothesis by conducting experimental word learning studies with adults using the paired-associate learning paradigm.

The second aim was to further investigate the PSTM hypothesis by examining whether an alternative learning paradigm, that of the Hebb repetition task, can be viewed as an experimental analogue of new phonological word-form learning. Chapter Two reviewed numerous studies which have investigated the factors that influence the Hebb Effect (e.g. Bower & Winzenz, 1969; Cohen & Johansson, 1967a, 1967b; Cunningham et al., 1984; Hebb, 1961; Hitch et al., 2005; McKelvie, 1987; Melton, 1963; Schwartz & Bryden, 1971; Sechler & Watkins, 1991). More recent studies have examined the effects of articulatory suppression and phonological similarity on the learning of sequences of familiar items such as digits, letters and words (e.g. Fallon et al., 2005; Hitch et al., in press; Page et al., 2006). These studies showed that articulatory suppression and phonological similarity failed to disrupt the learning of these sequences, further suggesting that the learning of familiar material is not mediated by PSTM. Such findings are also predicted by a computational model of PSTM for serial order. The Neural Network Model of the Phonological Loop (Burgess & Hitch, 1999, 2006) assumes that sequence learning occurs via connections between a context/timing signal and item representations, and as such predicts that sequence learning will not be disrupted by phonological manipulations. However, research examining the effect of phonological manipulations on the learning of sequences of unfamiliar material needs to be conducted. If the long-term learning of unfamiliar material is mediated by PSTM, then the learning of sequences of unfamiliar material should be disrupted by phonological manipulations, such as phonological similarity.

7.3 Summary of Main Empirical Findings

This section provides a summary of the main empirical findings. The results of the ISR and paired-associate experiments will be reported first (Experiments 1 to 3), followed by the results of the Hebb repetition experiments (Experiments 4 to 7).

7.3.1 Immediate Serial Recall and Paired-Associate Learning Experiments

Experiment 1 took the format of an ISR task. The nonwords selected had been previously rated as being low in wordlikeness. The results showed clear PSEs for both words and nonwords. Moreover, the size of the PSE was equivalent for both types of material. This confirmed that phonological similarity had been adequately manipulated. Evidence was also found for a lexicality effect (e.g. Hulme et al., 1991, 1995) as words were better recalled than nonwords, regardless of their phonological similarity.

In Experiment 2, the effect of phonological similarity on the learning of word pairs was only assessed at Trial 1 due to the near-ceiling performance observed for these pairs from Trial 2 onwards. No effect of phonological similarity on learning word pairs at Trial 1 was observed. In contrast, phonological similarity impaired the learning of nonword pairs when learning was assessed over all five trials; slower learning was observed for similar nonword pairs. However, when removing Trial 1 from the analysis due to floor effects for nonword pairs, the effect of phonological similarity on learning nonword pairs was eliminated. However, a further analysis based on learning gradients when excluding Trial 1 showed a marginally non-significantly higher learning gradient for distinct compared to similar nonword pairs.

Experiment 3 attempted to address these ceiling and floor effects by developing a novel matching procedure. This procedure involved extending nonword pair learning over 12 trials with the intention of selecting the nonword pair trial at which performance matched word pair performance at the start of learning. The effect of phonological similarity on the learning of wordlike nonword pairs was also investigated. The effect of phonological similarity on the learning of word pairs was restricted due to near-ceiling performance from Trial 3 onwards. The results based on word pair performance at Trials 1 and 2 revealed no effect of phonological similarity at either of these two trials. However, faster learning was observed for similar word pairs over Trials 1 and 2. In contrast, phonological similarity slowed down the learning of similar wordlike and unwordlike pairs when learning was extended over 12 trials. Moreover, the detrimental effect of phonological similarity appeared to be restricted to an 'intermediate' phase of learning for both wordlike and unwordlike pairs. Finally, the wordlike pairs were learned faster than the unwordlike pairs.

An analysis of between-trial learning and forgetting rates on the nonword pair results from Experiment 2, using an application of a Markov model, revealed that phonological similarity had its negative impact on between-trial forgetting rates only. Furthermore, error analyses revealed a predominance of phonological errors for similar nonwords in both Experiments 2 and 3. Experiment 3 showed that more of these errors were made for unwordlike nonword pairs at the intermediate phase of learning only.

In summary, Experiments 2 and 3 provided a partial replication of Papagno and Vallar (1992) and extended their findings to English participants and materials. In addition, Experiment 3 generated some novel findings which were not reported in Papagno and Vallar (1992) and Experiment 2. These will be discussed in more detail in section 7.4.1.

7.3.2 Hebb Repetition Experiments

Experiments 4 and 6 generated broadly parallel patterns of results. Firstly, evidence was found for reliable Hebb Effects for word and nonword sequences, although the Hebb Effect was somewhat weaker for nonword sequences in Experiment 4. In each case, cumulative learning over trials for Hebb sequences was shown to exceed any learning over trials for filler sequences. In addition, better overall performance was observed for Hebb than filler lists, with the exception of nonword sequences in Experiment 4, in which comparable overall levels of performance for Hebb and filler lists were reported. Secondly, a PSE was observed on ISR performance for both Hebb and filler word sequences. In contrast, PSEs were shown on ISR performance for Hebb but not filler nonword sequences. Thirdly, phonological similarity failed to affect the learning of word sequences, as evidenced by equivalent rates of learning the similar and distinct sequences. In contrast, phonological similarity disrupted the learning of nonword sequences, as significantly slower learning was observed for similar Hebb sequences. However, Experiment 6 also showed evidence of significant learning over trials for distinct filler nonword sequences. Analyses of between-trial learning and forgetting rates using a Markov model showed a converging pattern of results across Experiments 4 and 6. Phonological similarity affected learning and forgetting rates for word sequences. In contrast, phonological similarity had its negative impact only on forgetting rates for nonword sequences. Finally, simple error analyses showed that phonological similarity increased the propensity of order errors for both similar word and nonword sequences in Experiments 4 and 6. Hebb repetition increased the production of both total item errors and order errors for filler compared to Hebb word sequences in both Experiments 4 and 6. For nonword sequences, Hebb repetition only increased the propensity of total item errors for filler than Hebb lists in Experiment 6.

Experiments 5 and 7 produced a rather complex pattern of results that were of somewhat limited value due to the absence of reliable Hebb Effects for both word sequences (Experiment 5) and nonword sequences (Experiments 5 and 7). Firstly, weak evidence was shown for a Hebb Effect for word sequences in Experiment 5, as indicated by better overall performance for Hebb than filler sequences only. Furthermore, only poor evidence was found for Hebb Effects for nonword sequences in Experiments 5 and 7. In both experiments, although cumulative learning over trials was observed, this was comparable for Hebb and filler nonword sequences. Secondly, PSEs were shown on ISR performance for both Hebb and filler word and nonword sequences. Thirdly, phonological similarity failed to affect the learning of word sequences in Experiment 5, as shown by equivalent rates of learning the similar and distinct sequences. The effect of phonological similarity on the learning of nonword sequences was restricted due to the poor Hebb Effects in Experiments 5 and 7. The Markov model analyses of between-trial learning and forgetting rates were restricted due to the absence of reliable sequence learning over trials for both word and nonword sequences. Finally, simple error analyses showed that phonological similarity increased the production of order errors for both similar word sequences (Experiment 5) and similar nonword sequences (Experiments 5 and 7). Hebb repetition failed to affect the production of order errors for both word and nonword sequences (Experiments 5 and 7) and total item errors for nonword sequences (Experiment 5). Hebb repetition increased the production of total item errors for filler than Hebb lists for word sequences in Experiment 5.

A common finding across each of Experiments 4 to 7 was that nonword sequence learning was relatively poor, especially in comparison to the robust learning shown for nonword pairs in Papagno and Vallar (1992). Possible reasons for this finding will be discussed in section 7.5.

In summary, Experiments 4 and 6 provide converging evidence with paired-associate studies (e.g. Papagno & Vallar, 1992; Experiments 2 and 3). Experiments 5 and 7 are considered of limited value in terms of the effects of phonological similarity on nonword sequence learning, although they did highlight a number of methodological issues that need to be considered when conducting experiments using the Hebb repetition task. These issues will be discussed in sections 7.4.2 and 7.6.

7.4 Theoretical Interpretation of Findings

This section focuses on the theoretical interpretation of the empirical findings summarised in section 7.3. It will also consider the extent to which these findings provide evidence in support of the PSTM hypothesis. The theoretical interpretation of the paired-associate learning results will be discussed first, followed by the Hebb sequence learning results. The extent to which these two learning paradigms produce a converging pattern of results will be discussed in section 7.5.

7.4.1 The Role of Phonological Short-Term Memory in the Paired-Associate Learning of Word and Nonword Pairs

Given that previous studies have shown that PSTM mediates the paired-associate learning of unfamiliar material, but not the learning of familiar material (e.g. Papagno et al., 1991; Papagno & Vallar, 1992), Experiments 2 and 3 sought to further investigate this claim by examining the effects of phonological similarity on the learning of word-word and word-nonword pairs.

Word-word pair learning

The results of Experiments 2 and 3 provide limited evidence that PSTM is not utilised when learning familiar material. Phonological similarity failed to affect the learning of similar and distinct word pairs at Trial 1 (Experiments 2 and 3) and Trial 2 (Experiment 3). This suggests that participants are able to capitalise on existing lexical-semantic representations in LTM to learn word pairs, at least during the early stages of learning. As such, it is argued that participants rely primarily on non-phonological codes, such as semantic codes, to learn familiar material. This result provides a partial replication of Papagno and Vallar (1992) and extends their results to English participants and materials. An additional finding in Experiment 3 refers to the observation of the faster learning of similar nonword pairs over Trials 1 and 2; indeed, this result somewhat replicates Papagno and Vallar (1992).

However, a further aspect of Papagno and Vallar's (1992) word pair learning results was not replicated. Papagno and Vallar (1992) found a significant advantage for distinct word pairs at Trial 1, which they proposed reflected participants' use of PSTM during the initial stages of learning before shifting to alternative learning codes, such as semantic codes, to learn word pairs. Contrary to this finding, Experiments 2 and 3 reported no effect of phonological similarity on learning word pairs at Trial 1, suggesting that participants can adopt semantic strategies to learn these pairs from the outset of learning. Furthermore, the good level of performance observed for word pairs at Trial 1 in both Experiments 2 and 3 indicates that participants find it relatively easy to create semantic associations between words within a pair following only a single presentation. Moreover, this appears to remain the case even when attempting to reduce the ease with which semantic associations are generated by presenting more abstract cue words, as was the case in Experiment 3. The present findings may therefore imply that the paired-associate learning of word pairs does not necessarily require any contribution from PSTM.

However, there are two important points to note when interpreting these data. Firstly, given the ceiling effects observed for the learning of word pairs in Experiments 2 and 3, the interpretation of these results are arguably rather limited. This issue will be discussed further in section 7.6. Secondly, the reliability of Papagno and Vallar's (1992) findings, and subsequently their conclusions, may be restricted due to several methodological inadequacies within their experimental design (see Chapter Three). Nevertheless, despite these limitations, it is proposed that Experiments 2 and 3 provide limited evidence consistent with Papagno and Vallar (1992).

Word-nonword pair learning

The results of Experiment 2 provide limited evidence that PSTM mediates the learning of unfamiliar material. In line with Papagno and Vallar (1992), phonological similarity was shown to impair the learning of similar nonword pairs when learning was assessed over five trials. However, when performance at Trial 1 was excluded on the basis of a floor effect, no effect of phonological similarity on nonword pair learning was observed. This latter result suggests that phonological similarity does not affect the learning of unfamiliar material. However, a further analysis based on learning gradients when excluding Trial 1 provided some evidence that phonological similarity impaired the learning of similar nonword pairs, as a smaller learning gradient was observed for similar compared to distinct nonword pairs.

In contrast, the learning of nonword pairs in Experiment 3 generated a more complex and intriguing pattern of results when this learning was extended over 12 trials. Phonological similarity disrupted the learning of both wordlike and unwordlike pairs, although the evidence was somewhat weaker for the former type of pair; slower learning was observed for similar nonword pairs. As such, these results suggest that PSTM mediates the learning of unfamiliar material.

Taking the results of Experiments 2 and 3 together, this demonstrates that a variable known to affect the operation of the phonological loop component of the WM model (Baddeley & Hitch, 1974) in a clearly specified way (see Chapter One) reveals corresponding effects on the learning of nonword pairs, thus suggesting that PSTM is utilised when learning unfamiliar material. It is proposed that participants are forced into relying on PSTM to learn this material in the absence of existing lexical-semantic representations in LTM. As a result, participants create temporary representations based on the phonological structure of nonwords in order to construct more stable and long-lasting representations of these nonwords in LTM. This reliance on the phonological structure of nonwords is what elicits the effect of phonological similarity on learning. These findings are therefore consistent with Papagno and Vallar (1992), although the results of Experiment 2 are restricted by floor effects at Trial 1. Furthermore, it is important to bear in mind the questionable reliability of Papagno and Vallar's (1992) results due to their methodological drawbacks, as noted previously.

Extending the opportunity for learning nonword pairs and increasing task difficulty by presenting abstract cue words identified what appears to be three 'phases' of learning for both wordlike and unwordlike pairs. For both types of nonword pair, phonological similarity had a negative impact during an intermediate phase of learning.

This indicates that participants were utilising phonological coding, and hence PSTM, to learn the nonword pairs during this particular point in learning. However, the finding that phonological similarity failed to affect learning during the initial and latter trials suggests that either participants were not making use of PSTM to learn the nonword pairs or that PSTM is not the limiting factor when learning nonword pairs during these two particular phases. This finding cannot be easily attributed to floor and ceiling effects during the initial and latter phases, respectively, given that learning occurred during the initial phase and had not reached ceiling by the final trial.

There are a number of possible explanations for this pattern of results. Firstly, the absence of a significant effect of phonological similarity during the initial phase of learning may reflect participants' use of alternative non-phonological learning strategies. Another possibility is that participants found the task too difficult; that is, the task may initially overload PSTM capacity to the point where participants abandon the method of phonological coding. Indeed, in line with these points, previous research has suggested that participants abandon phonological coding in favour of alternative strategies when a task becomes too demanding (e.g. Baddeley, 2000b; Larsen & Baddeley, 2003; Larsen et al., 2000). Other research suggests that participants utilise strategies other than verbal rehearsal or may even switch strategies between trials (e.g. Logie et al., 1996). However, these findings are based on performance in the ISR task with familiar material and so therefore do not necessarily directly apply to the paired-associate learning of unfamiliar material. Further research needs to be conducted in order to determine which of these alternative explanations, if any, may account for the lack of a significant effect of phonological similarity during this initial learning phase. For example, an experiment could be conducted in which participants are instructed to utilise a verbal rehearsal strategy at each trial to ensure phonological coding within PSTM.

A further explanation for the pattern of learning observed during the initial learning phase may be related to the small number of similar pairs in PSTM at this phase. It may be the case that the fewer similar pairs there are in PSTM, the less chance there is of these pairs being confused. Indeed, this idea is consistent with Posner and Konick's (1966) 'acid bath' theory, whereby the rate of information loss from PSTM increases as the number of stored items in PSTM and their degree of phonological similarity increases. Hence, if only a few similar nonwords are in PSTM during the initial phase of learning, these may not be confused and, as a result, the learning of nonword pairs would not be impaired by phonological similarity. Further support for this idea may be shown by the finding that a comparable amount of similar nonwords were in PSTM for wordlike pairs (2.2 pairs) and unwordlike pairs (2.3 pairs) when the effect of phonological similarity first emerged.

In contrast, participants did appear to utilise phonological coding during an intermediate phase of learning. This may reflect the switching of strategies between the initial and intermediate phases of learning. Alternatively, it might indicate that the 'acid bath' becomes stronger; that is, as more similar nonwords are acquired, this increases the size of the pool of similar nonwords learned, which subsequently increases the level of confusability amongst these nonwords. In turn, this may slow down the learning of the similar nonword pairs.

Finally, the absence of a significant effect of phonological similarity during the latter phase of learning for both wordlike and unwordlike pairs may also reflect a number of possibilities. Firstly, participants may switch to using non-phonological learning strategies during this latter phase. On the other hand, it may be that the temporary representations of similar and distinct nonwords have become sufficiently stable at this point to enable these to become permanent representations in LTM. If this were the case, then the nonwords may no longer be susceptible to phonological similarity. A final possible explanation may be in terms of the observation of a 'glass ceiling'. The learning of distinct nonword pairs appears to reach an artificial asymptote during the latter phase of learning. Poorer learning over trials was observed for distinct compared to similar nonword pairs for both wordlike and unwordlike pairs during this particular phase of learning. It may be that this allows the learning of similar nonword pairs to 'catch up' with the learning of distinct nonword pairs. The identification of this glass ceiling for distinct nonword pairs is particularly interesting and possible reasons for its emergence are worth considering here. The glass ceiling may reflect differences in learning across participants. For example, some participants may succeed in learning all six distinct nonwords pairs, whereas other participants may fail to do so. Alternatively, it may be that the majority of participants learn all six distinct nonwords at some stage in learning, but then forget one or two pairs at a later stage. A further possible explanation for the occurrence of this glass ceiling may reflect the perseveration of errors; that is, participants may consistently make the same errors when attempting to learn the distinct nonword pairs which subsequently prevents the successful learning of all six nonword pairs. On the other hand, the emergence of a glass ceiling may simply be a feature of the paired-associate learning paradigm itself, although other studies using alternative tasks, such as nonword repetition, have also shown some evidence of a glass ceiling effect (e.g. S. E. Gathercole, personal communication, 2007; G. J. Hitch, personal communication, 2007).

The post-hoc explanations offered to account for the pattern of results obtained for nonword pair learning in Experiment 3 are somewhat speculative at this stage and further research is clearly needed in order to determine whether any of these explanations have any validity. Indeed, further experiments need to be conducted using the methodology adopted in Experiment 3 to confirm the presence of three phases of learning. However, the finding that phonological similarity had a similar effect on learning over trials for wordlike and unwordlike pairs may go some way towards supporting the idea of three phases of learning, although it is important to acknowledge that weaker evidence was found relating to the pattern of phonological similarity on wordlike pair learning.

The final theoretical interpretation of Experiment 3's results concerns a comparison of the learning observed for wordlike and unwordlike pairs. Wordlike pairs were learned faster over trials than unwordlike pairs. This finding converges with the results of Gathercole et al. (1996, cited in Gathercole & Martin, 1996) and suggests that the learning of wordlike pairs may make use of the existing structure of the language. Further evidence to suggest this refers to the 'location' of the intermediate phase of learning. The effect of phonological similarity emerged at Trial 3 for wordlike pairs and Trial 6 for unwordlike pairs, which may further suggest that the wordlike pairs were easier to learn.

However, despite these findings, it is clear that a relatively comparable pattern of results were obtained for wordlike and unwordlike pairs. This suggests that the two types of nonword pairs may be learned in a similar way and may rely on the same learning mechanisms. One possible reason for this finding may be related to the selection of the wordlike nonwords. Although the wordlike nonwords had significantly higher wordlikeness ratings, they were selected from a large cohort of nonwords that had been originally devised to be unwordlike. Hence, the wordlike nonwords may not be very wordlike. With this in mind, it would be useful to conduct Experiment 3 again but using a cohort of nonwords that were devised to more closely reflect the phonological structure of the language It would be interesting to determine whether a different pattern of wordlike pair learning would be found compared to that observed in Experiment 3.

Between-trial learning and forgetting of nonword pairs

The finding that phonological similarity had a detrimental effect only on between-trial forgetting rates for nonword pairs in Experiment 2 suggests that, although similar and distinct nonword pairs are equally difficult to learn, similar nonword pairs are particularly fragile and susceptible to forgetting between trials. That is, a similar nonword pair correctly learned on one trial is likely to be forgotten on the next trial. In contrast, the low between-trial forgetting rate observed for distinct nonword pairs suggests that when a distinct nonword pair is acquired on one trial, is it highly likely to remain intact on the next trial.

Detailed analyses of the types of errors made when learning similar and distinct nonword pairs in Experiment 2 offered an additional insight into the pattern of results generated from the Markov model analysis. The rarity of association errors suggests that similar nonword pairs are not forgotten between trials due to pairing a similar nonword with the incorrect cue word. Instead, similar nonword pairs are more likely to be forgotten between trials due to the fragile nature of individual similar nonwords. Indeed, the error analyses showed that more phonological errors were made for similar than distinct nonwords. This suggests that phonological representations of similar nonwords are particularly fragile. That is, it may be that the syllables and/or phonemes within a similar nonword are not bound together very tightly and are therefore more likely to fall apart between trials. In contrast, that distinct nonword pairings are more likely to remain intact between trials may suggest that the syllables and/or phonemes within distinct nonwords are bound together more tightly to form a whole nonword and are therefore less likely to fall apart between trials.

The pattern of between-trial learning and forgetting rates found in Experiment 2 could not be confirmed in Experiment 3 due to the elimination of a large proportion of participants' data. However, the error analyses conducted in Experiment 3 did confirm that more phonological errors were made for similar nonwords. Moreover, this finding was restricted to the intermediate phase of learning for unwordlike pairs. This parallels the pattern of correct recall performance for these nonwords, as phonological similarity was shown to impair learning during this particular phase of learning only.

To this end, it is suggested that the slower learning of similar nonword pairs over trials observed for correct recall performance is attributable to the fragility of similar nonwords between trials. In contrast, distinct nonwords appear more stable and as a result show better learning over trials in terms of correct recall performance.

Summary and conclusions

Experiments 2 and 3 provided a partial replication of Papagno and Vallar (1992) and extended their results to English participants and materials. PSTM was shown to mediate the learning of unfamiliar material in Experiment 3, and to a lesser degree in Experiment 2. Phonological similarity did not impair the learning of familiar material during the initial stages of learning (Trials 1 and 2) in Experiments 2 and 3. As such, it is proposed that these experiments provide some evidence to support the PSTM hypothesis, although it is important to note that Experiments 2 and 3 are somewhat limited to due ceiling- and floor effects for word (Experiments 2 and 3) and nonword (Experiment 2) pair learning. However, results from Experiment 3 generated novel insights into the learning of unfamiliar material, suggesting that additional processes may play a role in the long-term learning of new word-forms which may not be fully accounted for in terms of PSTM. Moreover, the Markov model analysis of between-trial learning and forgetting rates highlighted the role that forgetting plays in nonword pair learning and suggests that phonological representations of similar nonwords are particularly fragile.

7.4.2 The Role of Phonological Short-Term Memory in Word and Nonword Sequence Learning

Experiments 4, 5, 6 and 7 sought to replicate previous paired-associate studies (e.g. Papagno & Vallar, 1992). These experiments therefore investigated the effects of Hebb repetition and phonological similarity on word and nonword sequence learning.

Effect of Hebb repetition

Experiments 4 and 6, and to some degree Experiment 5, provided convincing evidence that learning a sequence of items can be facilitated by repeated order information. Evidence for clear and reliable Hebb Effects was found for word sequences in Experiments 4 and 6, and for nonword sequences in Experiment 6. Somewhat weaker Hebb Effects were also reported for word sequences in Experiment 5, and nonword sequences in Experiment 4. These findings are consistent with previous studies that have reported Hebb Effects for word sequences (e.g. Cumming et al., 2006; Gagnon et al., 2004; Page et al., 2006; Sechler & Watkins, 1991) and nonword sequences (e.g. Gagnon et al., 2004; Turcotte et al., 2005). Importantly, these findings confirm that the Hebb repetition paradigm can be used to investigate rates of learning familiar and unfamiliar material.

Effects of phonological similarity on ISR performance

The results of Experiments 4, 5, 6 and 7 suggest that participants utilise phonological coding during the immediate recall of word and nonword sequences. As such, this may provide tentative evidence that participants do not adopt non-phonological learning strategies, such as semantic strategies, when recalling individual sequences. In addition, there was some evidence in Experiments 4 and 6 to suggest that phonological similarity may not always be effective during the immediate recall of filler nonword sequences. In both experiments, PSEs were not obtained on ISR performance for these sequences. It is suggested that the emergence of a PSE on ISR performance may be related to the number of items correctly recalled within a sequence. That is, there need to be enough items correctly recalled for phonological similarity to have an effect on recall performance. Thus, given that only two nonwords were correctly recalled for filler nonword sequences in each of Experiments 4 and 6, phonological similarity failed to affect ISR performance. Indeed, the finding that PSEs were found on ISR performance for filler nonword sequences in Experiments 5 and 7 further confirms this idea, given that more than two nonwords were correctly recalled for these sequences in each of these experiments.

Effect of phonological similarity on sequence learning

The finding that phonological similarity failed to affect the learning of word sequences in Experiments 4, 5 and 6 provides support for the claim that PSTM does not mediate the learning of familiar material. It is proposed that participants can rely on existing lexical-semantic representations of words in LTM to learn the order in which a sequence of words is presented. Furthermore, the finding that PSEs were obtained on ISR performance for word sequences is presumably attributable to phonological similarity increasing the propensity of order errors for similar sequences (e.g. Conrad, 1965; Gathercole, et al., 2001; Henson et al., 1996). With this in mind, the finding that phonological similarity did not differentially affect the learning of similar and distinct word sequences suggests that learning the order of a sequence of similar words over repeated presentations is not necessarily hindered by the greater production of order errors for these sequences. Furthermore, the effect of phonological similarity on ISR performance during the course of learning suggests that learning repeated sequences does not diminish the role of phonological representations in immediate recall (Hitch et al., in press).

The pattern of word sequence learning results generated in Experiments 4, 5 and 6 are therefore consistent with previous studies which have examined the effect of phonological similarity on the learning of sequences of familiar material (e.g. Fallon et al., 2005; Hitch et al., in press; Page et al., 2006). Furthermore, they are in line with the predictions of the Burgess and Hitch (1999) model of the phonological loop. This model states that sequence learning arises through connections between a context/timing signal, which is responsible for the encoding of serial order, and item representations in an item layer. In contrast, phonological manipulations, such as phonological similarity, are dependent on connections between item and phoneme (input and output) layers and as such have their impact on the phonological/lexical layer of this model. Thus, the model predicts that learning sequences of familiar material will not be sensitive to phonological variables.

The effect of phonological similarity on nonword sequence learning was of particular interest given that this represented a novel area of investigation. Experiments 4 and 6 provided evidence that phonological similarity slowed down the learning of similar nonword sequences. This suggests that participants relied upon phonological coding to learn the nonword sequences. Given the absence of existing lexical-semantic representation of nonwords in LTM, participants are forced into relying on temporary representations based on the phonological structure of these nonwords in order to construct permanent and stable representations in LTM. As a result, the learning of these nonwords is susceptible to their phonological similarity. These results also provide a further demonstration that a variable known to have a specific effect on the operation of the phonological loop exerts a corresponding effect on the learning of unfamiliar material. Taking these findings together, it is proposed that the long-term learning of unfamiliar material appears to be mediated by PSTM. However, it is important to bear in mind that nonword sequence learning was relatively poor in both Experiments 4 and 6; this point will be expanded upon in section 7.5.

An additional conclusion may also be drawn from the nonword sequence learning results of Experiment 6. Evidence was provided to suggest that item learning can occur in the absence of repeated order information. Distinct filler sequences showed cumulative learning over trials. Indeed, this finding may have contributed to the absence of a significant three-way interaction between phonological similarity, list type and trials for nonword sequences in Experiment 6. This interaction was significant in Experiment 4 which suggests that distinct filler sequences failed to show cumulative learning over trials. It is not clear what may have caused this difference between the two experiments. It may be that participants found it easier to learn the distinct filler sequences in Experiment 6 given that the length of these were reduced down to five nonwords compared to six nonwords.

Between-trial learning and forgetting of word and nonword sequences

Experiments 4 and 6 generated consistent patterns of results in terms of between-trial learning and forgetting rates when using an application of a Markov model. The finding that phonological similarity affected both learning and forgetting rates for word sequences suggests that not only are similar word sequences harder to learn but that they are more unstable and susceptible to forgetting between trials than distinct word sequences. As such, it is proposed that learning the order of a sequence of words with similar phonological representations is particularly difficult, with similar words learned in the correct serial position on one trial being forgotten on the next trial. In contrast, once a distinct word has been learned in a particular position on one trial, it appears more likely to be retained on the next trial. Indeed, such findings fit in with the claim that phonological similarity impairs PSTM for order information by increasing the propensity of order errors (e.g. Conrad, 1965; Henson et al., 1996; Wickelgren, 1965b). Although the Markov model results in Experiment 5 provided further evidence that similar word sequences are vulnerable to forgetting between trials, these results are considered of limited value given that, based on the correct recall performance results, poor word sequence learning and a weak Hebb Effect were observed in this experiment.

For nonword sequences, phonological similarity was shown to selectively affect between-trial forgetting rates in Experiments 4 and 6, suggesting that although similar and distinct nonword sequences are equally difficult to learn, similar nonword sequences are particularly fragile and susceptible to forgetting between trials. It is proposed that not only are fewer similar nonwords recalled in the correct position on individual trials, but that similar nonwords are less likely to be recalled in the correct serial position between trials. In contrast, distinct nonwords learned in particular positions on one trial appear more likely to be retained on the next trial. As such, it is proposed that more distinct nonwords are learned and retained from trial to trial compared to similar nonwords. Although the Markov model results in Experiments 5 and 7 provided further evidence that similar nonword sequences are unstable and are likely to be forgotten between trials, the results are restricted due to the poor nonword sequence learning and poor Hebb Effects shown in both experiments based on correct recall performance.

The finding that phonological similarity affected between-trial forgetting rates in Experiments 4 and 6 may reflect different levels of fragility for word and nonword sequences. For example, similar word sequences may be fragile in terms of the ordering of the individual words within a sequence, given that words are already familiar. For similar nonword sequences, such fragility may operate at two levels: fragility may occur at the phonemic level of individual similar nonwords, given that nonwords do not have existing lexical-semantic representations in LTM, and/or at the level of the ordering of whole nonwords within a sequence. Difficulties at any one of these two levels may lead to impaired learning of similar nonword sequences. Indeed, it may be that the phonemes within similar nonwords are not bound together as tightly as they are for distinct nonwords; as a result, similar nonwords may be at a higher risk of falling apart between trials.

Error analyses for word and nonword sequences

Detailed error analyses conducted on word and nonword sequence learning in Experiments 4, 5, 6 and 7 presented a complex pattern of results which were difficult to interpret. As such, these analyses failed to shed further insights into the effects of phonological similarity and Hebb repetition on sequence learning. However, a series of more simple error analyses conducted on two main error types, total item errors and order errors, permitted an investigation into the overall effects of phonological similarity and Hebb repetition on sequence learning.

The error analyses conducted on word sequences revealed a consistent pattern of results across Experiments 4, 5 and 6 in terms of the effects of phonological similarity on the production of order errors. In each case, phonological similarity impaired the immediate serial recall of a sequence of similar words. This suggests that recalling a sequence of words is particularly difficult when that sequence contains similar rather than distinct words. In line with the results of the Markov model analyses for word sequences, these findings suggest that similar word sequences may be fragile in terms of the ordering of words within a sequence. Furthermore, these findings converge with the results of previous studies which have found that phonological similarity impairs the retention of order information (e.g. Conrad, 1965; Gathercole et al., 2001; Henson et al., 1996; Wickelgren, 1965b). The effect of phonological similarity on the production of total item errors was more varied across Experiments 4, 5 and 6, suggesting that participants erroneously recall words from outside the just-presented sequence for both similar and distinct word sequences. Finally, the finding that Hebb repetition had an effect on both total item errors and order errors in Experiments 4 and 6 may be related to the emergence of a reliable Hebb Effect based on correct recall performance. Indeed, Hebb repetition only affected total item errors in Experiment 5, in which only weak evidence for a Hebb Effect was observed.

For nonword sequences, phonological similarity was shown to impair the recall of an ordered sequence of similar nonwords across Experiments 4, 5, 6 and 7, as evidenced by a higher proportion of order errors for similar nonword sequences. This provides evidence that recalling a sequence of nonwords in the correct serial order is more difficult when that sequences contains similar nonwords. This finding also supports existing evidence that phonological similarity impairs PSTM for sequences of similar nonwords (e.g. Gathercole et al., 2001). Moreover, these findings may inform the results of the Markov model analyses to a certain degree; that is, the finding that phonological similarity increases the propensity of order errors for similar sequences suggests that these sequences are fragile at the level of the ordering of individual nonwords within a sequence. However, the current error analyses are not able to determine whether similar nonword sequences are also fragile at the level of phonemes within individual nonwords. As was noted for the word sequences, the effect of phonological similarity on total item errors was rather inconsistent over Experiments 4 to 7, suggesting that participants recall nonwords from outside the just-presented sequence for both similar and distinct sequences. Finally, the effect of Hebb repetition appears to be related to the emergence of a Hebb Effect based on correct recall performance. An effect of Hebb repetition was only observed in Experiment 6; this experiment was the only one in which a strong and reliable Hebb Effect on correct recall performance was reported.

Results with limited theoretical interpretation

The nonword sequence learning results of Experiments 5 and 7 are limited by the observation of poor Hebb Effects for these sequences, as mentioned in section 7.3.2. Given this situation, it is proposed that the effect of phonological similarity on nonword sequence learning cannot be reliably assessed in these two experiments.

Possible explanations for the absence of reliable Hebb Effects in Experiment 5 were discussed in detail in Chapter Six. To recap, it was proposed that the Hebb Effect is influenced by the degree of item overlap between the Hebb and filler lists. That is, reliable sequence learning may not be observed when the same set of items is presented for the Hebb and filler lists. It is suggested that this may reflect some kind of interference between the Hebb and filler lists such that the partial learning of the filler sequences may interfere with the learning of the Hebb sequences. A similar conclusion was drawn by Cumming et al. (2006). This idea may be supported by the presence of stronger Hebb Effects when presenting a different set of items for the Hebb list and each of the filler lists in Experiments 4 and 6.

The same explanation does not appear to account for the poor Hebb Effect observed in Experiment 7, given that there was no item overlap between the Hebb and filler lists. A possible explanation for this finding was detailed in Chapter Six. In brief, this suggests that the weak Hebb Effect may be due to the recall method of serial order reconstruction (SOR). It was tentatively proposed that in order to show cumulative learning over trials, participants may need to explicitly learn the phonological forms of nonwords; that is, some demand on item learning may be required. It may be that participants need to rely to a greater extent on the temporary phonological representations of nonwords in PSTM in order to benefit from repeated order information than is required when the method of recall is SOR. One possible way of testing this explanation would be to run the experiment with word sequences. The observation of a reliable Hebb Effect with word sequences may provide support for the idea that some item learning is required in order to learn nonword sequences. However, if no such Hebb Effect was reported, this may suggest that the findings of Experiment 7 may be related more directly to the SOR method of recall.

Finally, it is important to acknowledge that the poor Hebb Effects observed in Experiments 5 and 7 may simply be a problem of low experimental sensitivity. The particular designs adopted may not be powerful enough to detect any genuine effects of phonological similarity, or Hebb repetition, on learning nonword sequences. Increasing sample size may be one possible way of increasing experimental sensitivity.

Summary and conclusions

To conclude, it is tentatively proposed that Experiments 4 and 6 provide a converging pattern of results with paired-associate studies (e.g. Papagno & Vallar, 1992). That is, PSTM has been shown to support the learning of unfamiliar material, but not the learning of familiar material. To this end, Experiments 4 and 6 provide some evidence in support of the PSTM hypothesis. In contrast, the extent to which PSTM mediates the learning of unfamiliar could not be determined in Experiments 5 and 7 due to the lack of reliable nonword sequence learning and poor Hebb Effects.

7.4.3 Broader Implications of Findings for Theories and Computational Models of Phonological Short-Term Memory

The findings generated from the experimental work conducted in this thesis have a number of broader implications for theories and computational models of PSTM. Firstly, the findings obtained in Experiment 3 suggest that the role of PSTM in the learning of unfamiliar material may change over the course of learning. This may imply that additional mechanisms are available to support this learning other than PSTM. Various possible explanations were put forward in section 7.4.1 which may have implications for theories of PSTM. For example, the suggestion that participants may utilise strategies other than phonological coding when learning nonword pairs may indicate a role for long-term knowledge about the structure of language. Indeed, this supports previous research which has shown that the learning of unfamiliar material can be influenced by existing lexical knowledge of a language (e.g. Cheung, 1996; Gathercole et al., 1991a, 1999; Gathercole et al., 1996, cited in Gathercole & Martin, 1996; Roodenrys & Hinton, 2002; Thorn & Frankish, 2005; Thorn & Gathercole, 2001).

On a related note, the varying effect of phonological similarity on nonword pair learning observed in Experiment 3 is somewhat reminiscent of Gathercole et al.'s (1992) finding that the influence of PSTM in vocabulary acquisition appears to change during the course of development. To recap, PSTM skills were shown to influence vocabulary knowledge in children between the ages of 4 and 5 years. After this point in development, however, existing linguistic knowledge appears to influence PSTM skills. Gathercole et al. (1992) proposed that reliance on PSTM skills declines as vocabulary knowledge expands. This suggests that existing language knowledge plays an important role in vocabulary acquisition during the later stages of development. A similar conclusion may be made regarding the absence of an effect of phonological similarity during the latter phase of nonword pair learning observed in Experiment 3. This may reflect a reduction in the reliance on PSTM on the basis that participants' knowledge of the nonword pairs has increased by this latter phase. Although these findings are not directly comparable, the similarities are worth noting.

The experimental work conducted in Chapters Five and Six also has implications for computational models of PSTM which fail to implement mechanisms for the long-term learning of unfamiliar material (e.g. Brown et al., 2000; Hartley & Houghton, 1996; Henson, 1998; Page & Norris, 1998). In addition, the finding that phonological similarity disrupts nonword sequence learning (Experiments 4 and 6) presents a challenge for the Burgess and Hitch (1999) model of the phonological loop which postulates that sequence learning arises in the context/timing signal and is not affected by phonological variables. The current findings may therefore suggest that learning an ordered sequence of unfamiliar phonological representations, in this case nonwords, recruits the use of the phonological/lexical layer, as well as the context/timing signal. It may be that learning the order of phonemes within individual nonwords takes place in the phonological/lexical layer, with the context/timing signal being involved in learning the order of whole nonwords within a sequence. Whatever the process, the present results may go some way to providing constraints on the development of the Burgess and Hitch (1999) model to incorporate the long-term learning of a sequence of unfamiliar items.

7.5 Is the Hebb Repetition Paradigm an Analogue of New Word-Form Learning?

Given that an important aim of the current research was to determine whether the Hebb repetition paradigm could be viewed as an analogue to the long-term learning of new word-forms, it is important to consider the extent to which the results obtained using the Hebb repetition task paralleled those obtained using the paired-associate learning task.

Firstly, and perhaps most importantly, some evidence was found to suggest that PSTM mediates the learning of unfamiliar material, but not the learning of familiar material, in both the paired-associate (Experiments 2 and 3) and Hebb repetition (Experiments 4 and 6) tasks. As such, these results provide converging evidence in support of the PSTM hypothesis. Further support for a convergent pattern of results across the two learning paradigms concerns the finding that phonological similarity selectively affected between-trial forgetting rates for unfamiliar material when conducting a Markov model analysis (Experiments 2, 4 and 6). This correspondence appears to confirm that forgetting may make a large contribution to the patterns of nonword learning observed in these experiments. This is an important finding as the role of forgetting is rather neglected when assessing learning purely on the basis of correct recall performance results. However, a similar comparison cannot be made across the two learning tasks for the learning of familiar material as Markov model analyses of between-trial learning and forgetting rates were not conducted on the learning of word pairs in Experiments 2 and 3.

Despite the parallel patterns of results discussed above, there appear to be two important differences in the results generated from the two learning tasks. Firstly, the learning of familiar material in the two tasks appears to rely to a different extent on existing lexical representations. For example, it was proposed that the learning of word pairs in the paired-associate task does not rely on phonological coding and is instead presumably facilitated by the use of non-phonological codes, such as semantic codes (Experiments 2 and 3). This was tentatively shown by comparable levels of performance for similar and distinct word pairs at Trial 1 (Experiments 2 and 3) and Trial 2 (Experiment 3). In contrast, although equivalent rates of learning similar and distinct word sequences were found in the Hebb repetition task (Experiments 4, 5 and 6), ISR performance was impaired by phonological similarity, suggesting that participants do utilise phonological coding during this task, even if not when learning word sequences. These findings indicate that phonological representations of familiar items, such as words, have a role to play in the Hebb repetition task but not in the paired-associate task. It is therefore suggested that the learning of familiar material in the paired-associate task relies to a greater extent on the use of non-phonological learning codes which make use of existing lexical-semantic representations in LTM.

Perhaps the most striking difference between the two learning tasks is in terms of the degree of nonword learning observed. Robust nonword pair learning was observed in the paired-associate experiments; this is in line with previous paired-associate studies (e.g. Papagno et al., 1991; Papagno & Vallar, 1992). In contrast, nonword sequence learning was rather poor in each of the Hebb repetition experiments. Arguably, nonword learning is a two-stage process in both the paired-associate and Hebb repetition tasks. In the paired-associate task, learning a nonword pair involves learning the order of a novel sequence of phonemes (that is, a nonword) and learning to associate that nonword with a familiar lexical representation (i.e. the cue word). In the Hebb repetition task, learning a nonword sequence involves learning the order of a novel sequence of phonemes and learning the order of nonwords within a sequence. However, despite this claim of a two-stage learning process operating in both learning paradigms, it is tentatively suggested that the learning process involved in these two paradigms employs the use of different learning mechanisms.

It is proposed that nonword learning in the paired-associate task may be facilitated by the process of associating a nonword with a familiar lexical representation. It may be that this learning process benefits from the support of some sort of lexical 'hook' provided by the cue word. Conversely, nonword sequence learning in the Hebb repetition task does not appear to benefit from existing lexical representations given that no such lexical information is presented which may support this learning. With this in mind, it is tentatively proposed that the difference in the degree of nonword learning between the two paradigms may be related to the extent to which nonword learning can benefit from existing lexical knowledge of the language.

Taking all these points into consideration, a number of conclusions may be tentatively drawn. Firstly, the finding that two of the four Hebb repetition experiments conducted in this thesis (Experiments 4 and 6) provided a converging pattern of results with paired-associate studies (e.g. Papagno & Vallar, 1992; Experiments 2 and 3) provides limited evidence that the Hebb repetition paradigm may be a possible analogue of new word-form learning. However, it is argued that the observation of poor nonword sequence learning restricts the extent to which Experiments 4 and 6 provide conclusive evidence in support of this claim. It is clear that nonword sequence learning needs to be improved in future research in order to provide more conclusive support for idea that the Hebb repetition paradigm represents an analogue of new word-form learning. Possible ideas for improving nonword sequence learning will be discussed in section 7.7.

Secondly, despite the limited evidence that the Hebb repetition paradigm may possibly be used to investigate the role of PSTM in new word-form learning, it is tentatively proposed that the paired-associate learning paradigm may represent a closer analogue of 'real' word learning than the Hebb repetition paradigm, given that it makes use of existing lexical knowledge. This idea supports previous developmental research highlighting the interactive nature of PSTM and existing vocabulary knowledge (e.g. Gathercole et al., 1992). Indeed, the paired-associate paradigm may also be considered more like real word learning in terms of face validity than the Hebb repetition paradigm.

7.6 Methodological Issues

The current experiments have raised a number of important methodological issues which need to be acknowledged here. Two of these issues are specific to the particular learning paradigms used, whereas one applies more broadly to research investigating the learning of familiar and unfamiliar material. These methodological issues will be discussed in turn.

The first methodological issue that the current research has highlighted refers to the presence of ceiling and floor effects in Experiment 2 for word- and nonword pair learning, respectively. Such effects limit the statistical analysis of data and consequently the interpretation of the results. One possible way of eliminating ceiling and floor effects is to match performance across word and nonword pair learning at Trial 1. However, this appears to be difficult to accomplish given that the pilot experiments described in Chapter Four failed to satisfactorily match word- and nonword pair performance at Trial 1, despite implementing various design modifications. It is not clear why these pilot experiments failed to match word- and nonword pair performance. One possibility may relate to sample size; a maximum of ten participants took part in each of the pilots. It may be that the pilot experiments would have been more successful if a larger sample of participants had been recruited. Nevertheless, these pilot experiments served to highlight the methodological difficulties when attempting to match word- and nonword pair performance.

A form of matching performance was achievable by extending the opportunity for nonword pair learning in Experiment 3. This novel design involved selecting the nonword pair trial (Trial N) at which performance corresponded to the level of performance at Trial 1 for word pairs. However, it is important to note that even this matching procedure does not represent truly comparable levels of performance at Trial 1, given that the nonword pairs have undergone some learning prior to the selection of Trial N.

It is clear from Experiments 2 and 3 that the presence of ceiling and floor effects and the identification of a suitable matching methodology to reduce these effects present a challenge for research which aims to compare the learning of words and nonwords. However, presenting participants with the opportunity for extended nonword pair learning served to highlight a number of additional insights into how phonological similarity affects the learning of nonword pairs, as was discussed in detail in section 7.4.1.

A second methodological issue concerns the importance of obtaining strong and reliable Hebb Effects when investigating the effect of phonological similarity on learning during Hebb repetition tasks. The absence of reliable Hebb Effects for nonword sequences in Experiments 5 and 7 restricted the interpretation of these results. It was tentatively suggested that a number of design modifications may have influenced the emergence of a Hebb Effect, such as the degree of item overlap between the Hebb and filler lists (Experiment 5) and the method of recall adopted (Experiment 7). It is clear that careful consideration needs to be made when designing such experiments in order to promote the Hebb Effect. Future research would benefit from identifying other possible factors which may affect the Hebb Effect.

One broader methodological issue has also been identified. The present experiments have highlighted that simple analysis based on correct recall performance alone is not fully informative about the microstructure of learning. Although this method of analysis can be used to examine the effect of phonological similarity on the rate of learning familiar and unfamiliar material, it fails to provide information regarding the role that forgetting may play in this process. However, an analysis of between-trial learning and forgetting rates using an application of a Markov chain model permits a more detailed investigation into the stability of (i) word- and nonword pairings and (ii) word and nonword sequences. Indeed, this analysis has been applied to the majority of the experimental data reported in this thesis and has served to highlight additional insights concerning the effect of phonological similarity on paired-associate learning and sequence learning, as was discussed in section 7.4.

However, despite the benefits afforded by a Markov model analysis in terms of providing a more detailed examination of the process by which familiar and unfamiliar material is acquired, it is important to consider the negative aspects of conducting such an analysis. The current research has shown that the presence of ceiling and/or floor effects can severely restrict this analysis. For example, the ceiling effects shown for the learning of word pairs in Experiments 2 and 3 led to the elimination of the majority of participants' data, and the floor effects reported for nonword pair learning in Experiment 2 reduced the sample size by 39%. Furthermore, the near- ceiling and floor effects observed for extended nonword pair learning in Experiment 3 prevented a reliable analysis of between-trial learning and forgetting rates. As such, it is suggested that this method of analysis is only reliable when results avoid ceiling and floor effects.

A further point to consider when conducting a Markov model analysis concerns the problem of missing data; this problem can lead to a reduction in sample size. Such a problem was encountered in Experiments 4 to 7 and reflected instances in which (i) all items were learned on one trial and remain learned on the next trial and/or (ii) no items were learned on one trial with these items remaining unlearned on the next trial. However, a possible solution to this problem is to average transitional probabilities over all transitional steps. Although this procedure results in a loss of detailed information regarding the pattern of learning and forgetting rates over trials, it still allows insights to be gained regarding the effect of phonological similarity on sequence learning. However, when adopting this procedure it is important to confirm that the distribution of missing data is unbiased across transitional steps. This can be achieved by replacing missing transitional probabilities with individual transitional step means.

Finally, it is worth noting that Markov model analyses based on between-trial learning and forgetting rates may provide patterns of results which appear to contradict results based on correct recall performance. It is important to bear in mind that these two analyses represent different ways of investigating the data. A Markov model analysis of learning and forgetting rates takes into account whether the same items are recalled in the correct serial position between trials, whereas correct recall performance only takes into account the total number of items recalled in the correct serial position at each trial, regardless of whether these items are the same between trials.

After considering the advantages and disadvantages of the Markov model analysis, it is proposed that an analysis of between-trial learning and forgetting rates is a worthwhile investigation. The additional insights gained, particularly regarding the role of forgetting during learning, may provide the impetus for conducting future experimental investigations which are designed to test further predictions based on such findings. However, it is advised that future researchers intending to conduct this analysis should seek to reduce the possibility of ceiling and floor effects occurring during the learning process.

Finally, error analyses may also provide further information regarding the microstructure of learning which is not available from an analysis of correct recall performance alone. For example, error analyses allow a detailed examination of the stability of individual items and may highlight specific types of errors that contribute to patterns of learning and/or forgetting. However, error analyses may not be as informative when complex designs that manipulate a number of variables are adopted (e.g. Experiments 4 to 7) or when numbers of errors are too low (e.g. the learning of word pairs in Experiments 2 and 3).

7.7 Limitations and Ideas for Future Research

The experimental work reported in this thesis has suffered from a number of limitations. These limitations have been acknowledged and discussed at several points and the majority have been discussed in detail in the current chapter. Briefly, the main limitations refer to: (i) ceiling and floor effects for word- and nonword pair learning (Experiments 2 and 3); (ii) problems associated with developing a suitable matching procedure for word- and nonword pair learning (see Chapter Four); (iii) the lack of reliable Hebb Effects for word and nonword sequences (Experiments 5 and 7); and (iv) poor nonword sequence learning (Experiments 4, 5, 6 and 7). The current section focuses on offering ideas for future research, some of which attempt to provide solutions to some of the limitations encountered in this thesis.

First and foremost, it is essential that further research is conducted in order to confirm many aspects of the current work. For example, the proposal that extended nonword pair learning shows three phases of learning, and may therefore not rely exclusively on PSTM, needs further confirmation (Experiment 3). This finding would benefit from a replication that aims to extend these findings to different sets of materials. Perhaps an additional method of testing the findings of Experiment 3 would be to investigate the use of strategies during the learning process. This may shed further insights into whether participants switch strategies when learning the nonword pairs. Furthermore, the suggestion that wordlike and unwordlike nonwords show similar patterns of learning needs further clarification using nonwords that conform more closely to the concept of 'wordlike'. This may be achieved by selecting nonwords based on more objective measures such as phonotactic frequency and/or neighbourhood size.

Future work also needs to concentrate on determining the circumstances in which a Hebb Effect emerges. Two of the four experiments in this thesis reported only weak Hebb Effects (Experiments 5 and 7). Whilst this is an interesting finding in itself, it prevented a reliable examination of the effect of phonological similarity on nonword sequence learning. Chapter Two reviewed numerous studies which identified a number of factors influencing the Hebb Effect (e.g. Bower & Winzenz, 1969; Cohen & Johansson 1967a, 1967b; McKelvie, 1987; Melton, 1963); however, only a few of these studies used words as materials (e.g. Cumming et al., 2006; Page et al., 2006; Sechler & Watkins, 1991) with fewer still using nonwords as materials (e.g. Turcotte et al., 2005). It may be that factors which have been previously shown to promote a Hebb Effect are specific to the materials used. For example, Hebb (1961) found a Hebb Effect when using the same set of items (i.e. digits) as Hebb and filler lists. In contrast, a reliable Hebb Effect appears difficult to detect in experiments which adopt this design procedure using words and nonwords as the materials (e.g. Experiment 5; Cumming et al., 2006). Future research also needs to consider carefully the extent to which changes in the method of recall may have on the production of a Hebb Effect using different types of materials (see Experiment 7).

Another important line of future research would be to examine the effect of manipulating PSTM variables other than phonological similarity on the learning of familiar and unfamiliar material. Previous research has shown that additional PSTM variables, such as word length and articulatory suppression, also impair the learning of unfamiliar material, but not the learning of familiar material, when using the paired-associate task (e.g. Papagno et al., 1991; Papagno & Vallar, 1992). Such studies would benefit from replication in order to provide further support for the PSTM hypothesis. Indeed, it would be particularly interesting to examine whether PSTM variables such as word length and articulatory suppression have similar effects on extended nonword pair learning over trials as was shown in Experiment 3; that is, would three phases of learning emerge? Conducting such experiments may help to determine whether the absence of a significant effect of phonological similarity on nonword pair learning during the initial phase of learning reflects the idea that participants do not rely on phonological coding during these learning stages or whether an explanation in terms of an acid bath theory (e.g. Posner & Konick, 1966) accounts for this pattern of paired-associate learning.

Furthermore, it would also be of use to investigate the effects of PSTM variables such as word length and articulatory suppression on the learning of sequences of familiar and unfamiliar material in the Hebb repetition task. Recent research has already suggested that articulatory suppression does not impair sequence learning for familiar material (e.g. Fallon et al., 2005; Hitch et al., 2006; Page et al., 2006). It would be interesting to determine whether word length would have a corresponding effect on sequence learning for familiar material. More importantly, it would be of particular theoretical interest to investigate the effects of word length and articulatory suppression on the learning of sequences of unfamiliar material, such as nonwords. If these variables were shown to impair sequence learning for such material, this would not only generate further support for the PSTM hypothesis, but may provide more conclusive evidence that the Hebb repetition paradigm may be an analogue of new phonological word-form learning.

In order to provide reliable evidence that the Hebb repetition paradigm is an analogue of new word-form learning, a further important area for future research is to improve nonword sequence learning. The current work suggests that learning nonwords in the form of an ISR task does not promote robust sequence learning. One possible reason for this was discussed in section 7.5. An idea for future studies may be to use multisyllabic nonwords in place of a sequence of monosyllabic nonwords, as it may be that learning multisyllabic nonwords represents a more naturalistic new word-form learning environment. For example, instead of presenting a sequence of monosyllabic nonwords separated by brief intervals, these same nonwords could be concatenated to produce a single coarticulated multisyllabic nonword.

In line with this idea, research has shown that information which occurs in natural speech, such as prosody and coarticulation, improves nonword repetition accuracy (Archibald & Gathercole, in press; Nijland et al., 2002; Roy & Chiat, 2004). Archibald and Gathercole (in press) compared children's performance on a nonword serial recall task with their performance on a nonword repetition task, whilst matching the two tasks for phonological content. For example, consonant-vowel (CV) syllables were presented either in isolation (e.g. *fow...moy...chee*) or as a single coarticulated nonword (e.g. *fowmoychee*). Higher levels of repetition accuracy were found for the nonword repetition task, suggesting that additional cues, such as prosody and coarticulation, facilitate nonword repetition accuracy. Indeed, the authors concluded that "cues available in multisyllabic nonword repetition may allow for richer encoding ... resulting in better quality phonological representations that are less susceptible to interference or loss". Considering such findings, it may be feasible to suggest that nonword learning in the Hebb repetition task would be enhanced by the presentation of multisyllabic nonwords compared to sequences of monosyllabic nonwords. In turn, this may permit a more reliable and valid investigation as to whether the Hebb repetition paradigm represents an analogue to the long-term learning of new word-forms.

However, an important point to consider is the extent to which the recall of multisyllabic nonwords and sequences of monosyllabic nonwords rely on the same or different learning mechanisms. There is some evidence to suggest that common mechanisms underlie serial recall and nonword repetition tasks (e.g. Gupta, 2003, 2005). Gupta (2005) has reported standard primacy and recency effects in nonword repetition, proposing that a nonword may actually be processed as a sequence when it is first encountered. This finding may provide a further reason for investigating the learning of multisyllabic nonwords rather than sequences of monosyllabic nonwords in the Hebb repetition task.

Finally, a longer-term aim of the current research would be to extend the current findings to children. A few studies have used the paired-associate learning task with children to investigate the relationship between PSTM and vocabulary acquisition (e.g. Gathercole et al., 1997; Masoura & Gathercole, 2005). It would be interesting to conduct the current paired-associate experiments with children, although the manipulation of phonological similarity may preclude conducting such experiments with young children, given that they are less likely to be able to learn more than a few nonword pairs. Only one study appears to have examined sequence learning using the Hebb repetition task with children (e.g. Hitch et al., 2006) and this study used familiar items as the materials to be learned. If evidence is found with adults to support the idea that the Hebb repetition paradigm may possibly represent an analogue of new word-form learning, it would be of interest to investigate whether children are capable of showing a Hebb Effect for nonword sequences or indeed multisyllabic nonwords. Indeed, if phonological similarity was shown to selectively disrupt the learning of similar nonword sequences or multisyllabic nonwords, this would provide stronger evidence that the Hebb repetition paradigm is an analogue of new word-form learning and would provide support for the PSTM hypothesis. Moreover, given the finding that the influence of PSTM in vocabulary acquisition may change over development (e.g. Gathercole et al., 1992), it would be of particular interest to investigate whether learning in the Hebb repetition paradigm follows a similar developmental trajectory.

7.8 Conclusions

The experiments reported in this thesis examined the extent to which PSTM contributes to the long-term learning of new word-forms using paired-associate and Hebb repetition learning paradigms. Limited evidence was found to suggest that phonological similarity does not affect the learning of word pairs, at least during the initial stages of learning. In contrast, phonological similarity was shown to selectively disrupt the learning of nonword pairs, as evidenced by slower learning observed for similar nonword pairs. However, when providing the opportunity for extended learning, the detrimental effect of phonological similarity on nonword pair learning appears to be restricted to an intermediate phase of learning. Furthermore, this pattern of learning occurs for nonwords differing in their degree of wordlikeness. Taken together, these findings suggest that the learning of nonword pairs is mediated by PSTM, although the role of PSTM may change over the course of learning. This thesis also generated results which suggest that the Hebb repetition paradigm may be used to investigate rates of learning unfamiliar and familiar material. Reliable Hebb Effects were found for word and nonword sequences in two out of four Hebb repetition experiments. Phonological similarity was shown to impair ISR performance for word and nonword sequences, demonstrating that phonological coding is utilised during immediate recall of these sequences. Importantly, phonological similarity was shown to slow down the learning of nonword sequences, but not the learning of word sequences. This finding demonstrates that PSTM contributes to nonword sequence learning. Markov model analyses of between-trial learning and forgetting rates revealed that phonological similarity had its negative impact primarily on forgetting rates in both learning paradigms. Similar nonwords appear to be particularly fragile and susceptible to forgetting between trials. These findings serve to highlight the role of forgetting in the learning of unfamiliar material. Finally, the results of simple errors analyses indicate that phonological similarity impairs the immediate recall of similar word and nonword sequences. This suggests that recalling sequences of similar words or nonwords is particularly difficult.

In conclusion, both learning paradigms have demonstrated that PSTM mediates the learning of unfamiliar material, thereby providing support for the PSTM hypothesis. These findings are argued to provide limited evidence that the Hebb repetition paradigm may be a possible analogue to new word-form learning. However, future research needs to concentrate on improving nonword sequence learning in the Hebb repetition paradigm, as well as promoting reliable Hebb Effects, to further confirm and validate this claim. Finally, the paired-associate paradigm appears to benefit from existing lexical-semantic knowledge to a greater extent than the Hebb repetition paradigm, thereby representing a closer analogue of real vocabulary acquisition.

Appendix 1(a): Word Sets used in Experiment 1

	Similar Word Sets				Distinct Word Sets			
Materials Set	Word	K-F Frequency	Concreteness	No. of Syllables	Word	K-F Frequency	Concreteness	No. of Syllables
Materials A	Suspect	30	3.79	2	Narrow	63	3.72	2
	Supper	37	5.63	2	Lumber	35	5.60	2
	Sunset	14	5.25	2	Delight	29	2.82	2
	Hunter	18	5.35	2	Parish	11	4.34	2
	Hunger	17	4.10	2	Recall	39	3.19	2
	Budget	59	3.66	2	Mustard	20	5.95	2
	Bullet	28	5.95	2	Column	71	5.20	2
	Burden	44	3.91	2	Bishop	18	5.87	2
	Mean	31	4.71	2	Mean	36	4.59	2
Materials B	Pocket	46	5.78	2	Pupil	20	5.70	2
	Pollen	11	5.84	2	Rubber	15	5.96	2
	Powder	28	5.13	2	Safety	47	2.32	2
	Copper	13	5.47	2	Custom	14	3.23	2
	Content	53	3.00	2	Mortar	11	5.48	2
	Concert	39	2.52	2	Wedding	32	5.09	2
	Wonder	67	3.05	2	Highway	40	5.75	2
	Worker	30	5.32	2	Fellow	63	5.02	2
	Mean	36	4.51	2	Mean	30	4.82	2

Appendix 1(b): Nonword Sets used in Experiment 1

	Similar Nonword Sets			Distinct Nonword Sets		
Materials Set	Nonword	Wordlikeness Rating	No. of Syllables	Nonword	Wordlikeness Rating	No. of Syllables
Materials A	Merglip	1.35	2	Bamich	2.52	2
	Mefflib	1.39	2	Lebbist	2.26	2
	Memblin	2.78	2	Tafflost	2.17	2
	Feppip	1.61	2	Musglent	1.39	2
	Feggin	2.74	2	Cuddow	1.70	2
	Febslib	1.17	2	Pevtong	1.35	2
	Bebbict	1.96	2	Dapeth	2.09	2
	Berpict	2.43	2	Suttic	2.91	2
	Mean	1.93	2	Mean	2.05	2
Materials B	Paddip	2.70	2	Cepfil	2.17	2
	Paglip	1.96	2	Butkels	1.61	2
	Darglit	2.0	2	Jorlam	2.57	2
	Dasklint	1.65	2	Sibbart	2.39	2
	Damklin	2.30	2	Rodgunt	2.00	2
	Lappish	2.39	2	Welptar	1.83	2
	Larmip	2.22	2	Fiddop	3.30	2
	labblin	2.57	2	Ludgash	2.43	2
	Mean	2.22	2	Mean	2.29	2

Appendix 2(a): Cue Word Sets used in Experiment 2

Group	Word	K-F Frequency	Concreteness	No. of Syllables	Group	Word	K-F Frequency	Concreteness	No. of Syllables
Group 1	Anchor	15	5.95	2	Group 2	Arrow	14	5.95	2
	Escape	65	3.41	2		Duty	61	3.22	2
	Iron	43	5.84	2		Garden	60	6.02	2
	Mountain	33	6.16	2		Thunder	14	5.47	2
	Cottage	19	5.93	2		Football	36	5.84	2
	Giant	23	5.15	2		Navy	37	4.72	2
	Meadow	17	5.94	2		Talent	40	2.90	2
	Tractor	24	5.90	2		Weapon	42	5.60	2
	Mean	30	5.54	2		Mean	38	4.97	2
Group 3	Baby	62	5.89	2	Group 4	Butter	27	6.18	2
	Finish	39	3.43	2		Honey	25	6.11	2
	Needle	15	6.08	2		Lion	17	6.27	2
	Quarrel	20	3.79	2		Tennis	15	5.74	2
	Infant	11	5.79	2		Artist	57	5.54	2
	Lemon	18	6.08	2		Flower	23	5.84	2
	Palace	38	5.79	2		Repair	20	3.94	2
	Temper	12	3.53	2		Shadow	36	4.57	2
	Mean	27	5.05	2		Mean	28	5.52	2

Appendix 2(a): Cue Word Sets used in Experiment 2 (continued)

Group	Word	K-F Frequency	Concreteness	No. of Syllables	Group	Word	K-F Frequency	Concreteness	No. of Syllables
Group 5	China	69	5.78	2	Group 6	Damage	33	4.06	2
	Debate	32	3.75	2		Empty	64	3.74	2
	Obscure	17	3.20	2		Jacket	33	6.35	2
	Saddle	25	6.03	2		Upset	14	2.82	2
	Busy	58	3.29	2		Angel	18	3.99	2
	Kingdom	26	3.92	2		Fashion	69	3.56	2
	Object	65	4.87	2		Profit	28	3.64	2
	Vary	34	2.58	2		Sugar	34	6.20	2
	Mean	41	4.12	2		Mean	37	4.30	2
Group 7	Ferry	11	5.80	2	Group 8	Impact	67	3.79	2
	Mercy	20	2.39	2		Laughter	22	4.11	2
	Pencil	34	6.17	2		Rabbit	11	6.35	2
	Rescue	15	3.73	2		Sister	38	5.75	2
	Basket	17	6.06	2		Mirror	27	6.05	2
	Diet	21	4.49	2		Item	54	4.36	2
	Heaven	43	3.05	2		Reward	15	3.96	2
	Yellow	55	5.37	2		Vision	56	3.95	2
	Mean	27	4.63	2		Mean	36	4.79	2

Appendix 2(b): Cue-Target Pairs used in Experiment 2

Similar Word Sets		Distinct Word Sets	
Materials A	Materials B	Materials A	Materials B
Baby - Suspect	Butter – Pocket	Anchor – Narrow	Arrow – Pupil
Finish – Hunter	Honey – Copper	Escape – Lumber	Duty – Rubber
Needle – Supper	Lion – Wonder	Iron – Delight	Garden – Safety
Quarrel – Hunger	Tennis – Worker	Mountain – Parish	Thunder – Custom
Infant – Budget	Artist – Pollen	Cottage – Recall	Football – Mortar
Lemon – Bullet	Flower – Content	Giant – Mustard	Navy – Wedding
Palace – Burden	Repair – Concert	Meadow – Column	Talent – Highway
Temper - Sunset	Shadow – Powder	Tractor - Bishop	Weapon – Fellow

Similar Nonword Sets		Distinct Nonword Sets	
Materials A	Materials B	Materials A	Materials B
Rescue – Merglip	Sister – Paddip	Saddle – Bamich	Damage – Cepfil
Mercy – Feppip	Mirror – Darglit	Busy – Lebbist	Sugar – Butkels
Ferry – Bebbict	Reward – Dasklint	Debate - Tafflost	Profit – Jorlam
Basket – Feggin	Item – Paglip	Obscure – Musglent	Fashion – Sibbart
Pencil – Febslib	Vision – Lappish	Object – Cuddow	Empty – Rodgunt
Diet – Mefflib	Impact – Larmip	China – Pevtong	Jacket - Welptar
Heaven – Berpict	Rabbit – Labblin	Kingdom – Dapeth	Angel – Fiddop
Yellow – Memblin	Laughter – Damklin	Vary – Suttic	Upset – Ludgash

Appendix 2(c): Formulae used to Calculate Markov Model Transitional Probabilities

The following constitutes an example of the procedure followed to calculate the transitional probabilitiy:

- That an item in an unlearned state on one trial remains unlearned on the next trial (a)
- **That an item in an unlearned state on one trial is learned on the next trial (b)**
- That an item in a learned state on one trial remains learned on the next trial (c)
- **That an item in a learned state on one trial is forgotten on the next trial (d)**

If a participant's response on three trials was:

Serial Position	Trial n	Trial n+1	Trial n+2
1	√	X	√
2	√	√	X
3	X	√	X
4	X	√	√
5	√	X	√
Number Correct	3	3	3

Then the formulae for the transitional probabilities (a), (b), (c) and (d) would be:

Transitional Step Trial n to Trial n+1	Transitional Step Trial n+1 to Trial n+2
(a) $X_{n+1} / X_n \longrightarrow 0/2$	(a) $X_{n+2} / X_{n+1} \longrightarrow 0/2$
(b) $\surd_{n+1} / X_n \longrightarrow 2/2$	(b) $\surd_{n+2} / X_{n+1} \longrightarrow 2/2$
(c) $\surd_{n+1} / \surd_n \longrightarrow 1/3$	(c) $\surd_{n+2} / \surd_{n+1} \longrightarrow 1/3$
(d) $X_{n+1} / \surd_n \longrightarrow 2/3$	(d) $X_{n+2} / \surd_{n+1} \longrightarrow 2/3$

Where: $_n$ represents the Trial number,
X represents an incorrect response and
√ represents a correct response

Appendix 2(d): Error Proportions for Word-Word Pairs in Experiment 2

Table 2(d): *Mean proportion (and standard deviations) of the three error types obtained at each trial for similar and distinct word-word conditions*

	Condition/Trial											
	Similar Word Pairs						Distinct Word Pairs					
Type of Error	Trial 1	Trial 2	Trial 3	Trial 4	Trial 5	Mean (SD)	Trial 1	Trial 2	Trial 3	Trial 4	Trial 5	Mean (SD)
Omissions	.250 (.206)	.049 (.098)	.011 (.052)	.005 (.026)	-	.063 (.066)	.190 (.199)	.049 (.098)	.016 (.043)	-	-	.051 (.059)
Association	.011 (.036)	-	-	-	-	.002 (.007)	.022 (.081)	-	.005 (.026)	-	-	.005 (.005)
Item	.049 (.082)	.022 (.048)	.022 (.048)	.005 (.026)	-	.020 (.026)	.027 (.053)	.011 (.036)	.005 (.026)	-	-	.009 (.014)

Appendix 2(e): Error Proportions for Word-Nonword Pairs in Experiment 2

Table 2(e): *Mean proportion (and standard deviations) for the three error types at each trial for similar and distinct word-nonword conditions*

	Condition/Trial											
	Similar Nonwords						Distinct Nonwords					
Type of Error	Trial 1	Trial 2	Trial 3	Trial 4	Trial 5	*Mean (SD)*	Trial 1	Trial 2	Trial 3	Trial 4	Trial 5	*Mean (SD)*
Omissions	.609	.370	.288	.212	.163	.328	.674	.364	.239	.158	.082	.303
	(.233)	(.231)	(.048)	(.194)	(.197)	(.190)	(.199)	(.232)	(.269)	(.197)	(.154)	(.191)
Association	.022	.016	.005	.016	.011	.014	.011	.022	.011	.005	.011	.012
	(.048)	(.057)	(.026)	(.043)	(.036)	(.024)	(.036)	(.048)	(.036)	(.026)	(.036)	(.017)
Item	.326	.408	.364	.326	.299	.345	.239	.288	.255	.179	.190	.230
	(.252)	(.259)	(.232)	(.228)	(.241)	(.193)	(.164)	(.208)	(.268)	(.180)	(.219)	(.158)

Appendix 3(a): Cue Word Sets used in Experiment 3

Group	Word	K-F Frequency	Concreteness	No. of Syllables	Group	Word	K-F Frequency	Concreteness	No. of Syllables
Group 1	Essence	15	2.43	2	Group 2	Abrupt	18	2.66	2
	Finish	39	3.43	2		Childhood	50	3.35	2
	Glory	21	3.04	2		Clever	17	3.13	2
	Passion	28	3.00	2		Magic	37	2.57	2
	Context	35	3.14	2		Recruit	10	3.93	2
	Temper	12	3.53	2		Defeat	31	3.63	2
	Mean	25	3.10	2		Mean	27	3.21	2
Group 3	Dispute	34	3.50	2	Group 4	Devil	25	2.74	2
	Sorrow	9	2.82	2		Humble	18	2.31	2
	Neglect	12	2.82	2		Shallow	14	3.30	2
	Severe	39	2.86	2		Retain	11	3.08	2
	Eager	27	3.02	2		Talent	40	2.90	2
	Mischief	5	3.25	2		Predict	8	2.88	2
	Mean	21	3.05			Mean	19	2.87	2
Group 5	Condemn	4	3.14	2	Group 6	Kingdom	26	3.92	2
	Mercy	20	2.39	2		Debate	32	3.75	2
	Polite	7	3.42	2		Obscure	17	3.20	2
	Rescue	15	3.73	2		Restore	9	2.75	2
	Extra	50	2.62	2		Busy	58	3.29	2
	Heaven	43	3.05	2		Farewell	14	3.61	2
	Mean	23	3.06	2		Mean	26	3.42	2

Appendix 3(a): Cue Word Sets used in Experiment 3 (continued)

Group	Word	K-F Frequency	Concreteness	No. of Syllables	Group	Word	K-F Frequency	Concreteness	No. of Syllables
Group 7	Assist	26	3.42	2	Group 8	Boredom	11	2.62	2
	Reward	15	3.96	2		Disgrace	3	2.78	2
	Marvel	6	2.93	2		Impulse	20	2.71	2
	Challenge	36	3.08	2		Anger	48	3.15	2
	Insight	22	2.70	2		Profit	28	3.64	2
	Weakness	46	2.57	2		Forecast	10	3.08	2
	Mean	25	3.11	2		Mean	20	3.00	2
Group 9	Distinct	42	2.47	2	Group 10	Despair	21	2.79	2
	Idle	13	3.18	2		Cheerful	10	3.62	2
	Courage	32	2.77	2		Greedy	5	3.60	2
	Admire	10	2.96	2		Advice	51	2.91	2
	Friendship	27	3.35	2		Merit	29	3.08	2
	Horror	17	3.41	2		Hazard	12	3.42	2
	Mean	24	3.02	2		Mean	21	3.24	2
Group 11	Motive	22	2.55	2	Group 12	Fortune	25	3.64	2
	Revenge	7	3.05	2		Panic	22	3.24	2
	Blessing	10	2.77	2		Kindness	5	3.61	2
	Tribute	24	3.39	2		Intense	40	3.61	2
	Culture	58	3.51	2		Hatred	20	2.39	2
	Finance	31	3.71	2		Deceit	2	2.57	2
	Mean	25	3.16	2		Mean	19	3.18	2

Appendix 3(b): Target Word Sets used in Experiment 3

	Similar Word Sets				Distinct Word Sets			
Materials Set	Word	K-F Frequency	Concreteness	No. of Syllables	Word	K-F Frequency	Concreteness	No. of Syllables
Materials A	Suspect	30	3.79	2	Narrow	63	3.72	2
	Sunset	14	5.25	2	Lumber	35	5.60	2
	Hunter	18	5.35	2	Parish	11	4.34	2
	Hunger	17	4.10	2	Recall	39	3.19	2
	Budget	59	3.66	2	Mustard	20	5.95	2
	Burden	44	3.91	2	Bishop	18	5.87	2
	Mean	30	4.34	2	Mean	31	4.78	2
Materials B	Pollen	11	5.84	2	Rubber	15	5.96	2
	Powder	28	5.13	2	Safety	47	2.32	2
	Copper	13	5.47	2	Custom	14	3.23	2
	Content	53	3.00	2	Wedding	32	5.09	2
	Wonder	67	3.05	2	Highway	40	5.75	2
	Worker	30	5.32	2	Fellow	63	5.02	2
	Mean	34	4.64	2	Mean	35	4.71	2

Appendix 3(c): Target Unwordlike Nonword Sets used in Experiment 3

	Similar Unwordlike Nonword Sets			Distinct Unwordlike Nonword Sets		
Materials Set	Nonword	Wordlikeness Rating	No. of Syllables	Nonword	Wordlikeness Rating	No. of Syllables
Materials A	Merglip	1.35	2	Bamich	2.52	2
	Mefflib	1.39	2	Lebbist	2.26	2
	Feppip	1.61	2	Tafflost	2.17	2
	Feggin	2.74	2	Musglent	1.39	2
	Bebbict	1.96	2	Cuddow	1.70	2
	Berpict	2.43	2	Dapeth	2.09	2
	Mean	1.91	2	Mean	2.02	2
Materials B	Paddip	2.70	2	Cepfil	2.17	2
	Paglip	1.96	2	Butkels	1.61	2
	Dasklint	1.65	2	Jorlam	2.57	2
	Damklin	2.30	2	Sibbart	2.39	2
	Lappish	2.39	2	Rodgunt	2.00	2
	labblin	2.57	2	Welptar	1.83	2
	Mean	2.26	2	Mean	2.10	2

Appendix 3(d): Target Wordlike Nonword Sets used in Experiment 3

	Similar Wordlike Nonword Sets			Distinct Wordlike Nonword Sets		
Materials Set	Nonword	Wordlikeness Rating	No. of Syllables	Nonword	Wordlikeness Rating	No. of Syllables
Materials A	Meppict	2.61	2	Putchel	3.09	2
	Mellib	2.39	2	Fabbor	3.00	2
	Pedmin	2.91	2	Turlict	2.39	2
	Pefflin	2.87	2	Sappesh	1.91	2
	Sellict	3.26	2	Bergops	2.48	2
	Sempib	1.78	2	Darpist	2.87	2
	Mean	2.64	2	Mean	2.62	2
Materials B	Parvit	2.78	2	Webbist	3.17	2
	Pattish	3.43	2	Mordast	3.04	2
	Lappint	2.65	2	Purldam	3.00	2
	Lattip	2.70	2	Higgart	2.61	2
	Rasbit	3.00	2	Roskurl	2.48	2
	Raftip	3.22	2	Vernash	3.13	2
	Mean	2.96	2	Mean	2.91	2

Appendix 3(e): Cue-Target Pairs used in Experiment 3

Similar Word Pairs		Distinct Word Pairs	
Materials A	Materials B	Materials A	Materials B
Essence – Suspect	Dispute – Powder	Abrupt – Narrow	Devil – Rubber
Finish – Hunter	Sorrow – Copper	Childhood – Lumber	Humble – Safety
Glory – Hunger	Neglect – Wonder	Clever – Parish	Shallow – Custom
Passion – Budget	Severe – Worker	Magic – Recall	Retain – Wedding
Context – Burden	Eager – Pollen	Recruit – Mustard	Talent – Highway
Temper – Sunset	Content – Mischief	Defeat – Bishop	Predict – Fellow

Similar Unwordlike Nonword Pairs		Distinct Unwordlike Nonword Pairs	
Materials A	Materials B	Materials A	Materials B
Condemn – Bebbict	Assist – Labblin	Kingdom – Dapeth	Boredom – Cepfil
Mercy – Feppip	Reward – Dasklint	Debate – Tafflost	Disgrace – Rodgunt
Polite – Feggin	Marvel – Lappish	Obscure – Musglent	Impulse – Welptar
Rescue – Merglip	Challenge – Paddip	Restore – Bamich	Anger – Sibbart
Extra – Mefflib	Insight – Damklin	Busy – Lebbist	Profit – Jorlam
Heaven – Berpict	Weakness – Paglip	Farewell – Cuddow	Forecast – Butkels

Similar Wordlike Nonword Pairs		Distinct Wordlike Nonword Pairs	
Materials A	Materials B	Materials A	Materials B
Distinct – Meppict	Motive – Lappint	Despair – Putchel	Fortune – Webbist
Idle – Mellib	Revenge – Parvit	Cheerful – Fabbor	Panic – Mordast
Courage – Pedmin	Blessing – Rasbit	Greedy – Turlict	Kindness – Purldam
Admire – Peflin	Tribute – Pattish	Advice – Sappesh	Intense – Higgart
Friendship – Sellict	Culture – Raftip	Merit – Bergops	Hatred – Roskurl
Horror – Sempib	Finance – Lattip	Hazard – Darpist	Deceit – Vernash

Appendix 3(f): Error Proportions for Word-Word Pairs in Experiment 3

Table 3(f): *Mean proportions (and standard deviations) for the three error types obtained at each trial for similar and distinct word-word conditions.*

	Condition/Trial											
	Similar Words						Distinct Words					
Type of Error	Trial 1	Trial 2	Trial 3	Trial 4	Trial 5	*Mean (SD)*	Trial 1	Trial 2	Trial 3	Trial 4	Trial 5	*Mean (SD)*
Omissions	.375	.125	.052	.063	.010	.125	.354	.125	.031	.042	.010	.113
	(.177)	(.155)	(.169)	(.120)	(.042)	(.076)	(.210)	(.129)	(.067)	(.096)	(.042)	(.085)
Association	.104	.021	.010	.010	-	.029	.052	.052	.021	.010	-	.027
	(148)	(.057)	(.042)	(.042)		(.044)	(.100)	(.100)	(.057)	(.042)		(.046)
Item	.136	.031	.031	.010	.042	.050	.073	.052	.042	.021	.021	.042
	(.109)	(.067)	(.067)	(.042)	(.075)	(.040)	(.121)	(.117)	(.075)	(.057)	(.057)	(.063)

Appendix 3(g): Error Proportions for Wordlike Pairs in Experiment 3

Table 3(g): *Mean proportions (and standard deviations) for the three error types at each trial for similar and distinct word-wordlike conditions*

	Trial												
Condition/Error Type	1	2	3	4	5	6	7	8	9	10	11	12	Mean (SD)
Similar Wordlike Nonwords													
Omissions	.646 (.257)	.458 (.240)	.281 (.233)	.271 (.201)	.188 (.210)	.146 (.191)	.094 (.136)	.052 (.100)	.063 (.120)	.052 (.100)	.031 (.091)	.031 (.067)	.193 (114)
Association	.021 (.057)	.052 (.100)	.063 (.103)	.052 (.100)	.021 (.083)	-	.021 (.083)	-	.031 (.091)	-	-	-	.022 (.023)
Item	.250 (.243)	.281 (.249)	.292 (.269)	.260 (.243)	.313 (.257)	.281 (.263)	.281 (.284)	.250 (.298)	.167 (.211)	.188 (.235)	.208 (.232)	.156 (.206)	.244 (.175)
Distinct Wordlike Nonwords													
Omissions	.677 (.166)	.396 (.181)	.229 (.171)	.094 (.136)	.083 (.136)	.052 (.100)	.042 (.075)	.031 (.067)	.010 (.042)	-	.031 (.091)	.010 (.042)	.138 (.073)
Association	-	.042 (.075)	-	.021 (.057)	.010 (.042)	.010 (.042)	.031 (.067)	.010 (.042)	-	-	.010 (.042)	.010 (.042)	.012 (.026)
Item	.219 (.158)	.250 (.136)	.208 (.197)	.219 (.180)	.177 (.177)	.156 (.166)	.135 (.163)	.125 (.177)	.115 (.208)	.156 (.223)	.094 (.161)	.104 (.181)	.163 (.142)

Appendix 3(h): Error Proportions for Unwordlike Pairs in Experiment 3

Table 3(h): *Mean proportions (and standard deviations) for the three error types when collapsing trials into three learning phases for similar and distinct unwordlike nonword conditions*

	Condition/Learning Phase							
	Similar Unwordlike Nonwords				Distinct Unwordlike Nonwords			
Type of Error	Phase 1 (T1-T5)	Phase 2 (T6-T9)	Phase 3 (T10-T12)	*Mean (SD)*	Phase 1 (T1-T5)	Phase 2 (T6-T9)	Phase 3 (T10-T12)	*Mean (SD)*
Omissions	.284	.073	.024	.127	.308	.052	.028	.129
	(.107)	(.078)	(.035)	(.059)	(.142)	(.088)	(.073)	(.088)
Association	.031	.029	.024	.028	.031	.003	.007	.014
	(.033)	(.074)	(.057)	(.043)	(.069)	(.010)	(.019)	(.025)
Item	.444	.409	.198	.350	.417	.222	.156	.265
	(.200)	(.225)	(.216)	(.194)	(.161)	(.253)	(.234)	(.191)

Appendix 4(a): Word Sets used in Experiment 4

	Similar Word Sets								
Materials Set	Filler 1			Filler 2			Hebb		
	Word	K-F Frequency	Neighbourhood Size	Word	K-F Frequency	Neighbourhood Size	Word	K-F Frequency	Neighbourhood Size
Materials A	Lab	3	24	Tab	1	17	Rag	10	27
	Lad	6	34	Tag	5	24	Rat	6	37
	Lag	3	30	Tap	18	29	Rap	2	34
	Lap	19	33	Tan	9	28	Ram	2	35
	Mag	1	28	Pap	1	33	Cab	12	23
	Map	13	31	Pan	16	26	Cat	23	35
	Mat	5	38	Pad	8	25	Cam	1	30
	Mean	7.1	31.1	Mean	8.3	26.0	Mean	8.0	31.6
Materials B	Bud	9	31	Rub	6	23	Hub	11	21
	Bug	4	31	Rug	13	26	Hug	3	24
	Bus	34	24	Rum	3	33	Hum	5	28
	Bum	7	30	Rut	1	36	Hut	13	33
	Dud	1	22	Sub	5	23	Mud	32	25
	Dug	15	27	Sum	45	28	Mug	1	24
	Dun	1	41	Sup	1	21	Mum	1	27
	Mean	10.1	29.4	Mean	10.6	27.1	Mean	9.4	26.0

Appendix 4(a): Word Sets used in Experiment 4 (continued)

Materials Set	Distinct Word Sets								
	Filler 1			Filler 2			Hebb		
	Word	K-F Frequency	Neighbourhood Size	Word	K-F Frequency	Neighbourhood Size	Word	K-F Frequency	Neighbourhood Size
Materials A	Vet	1	21	Nut	15	33	Rod	18	32
	Tip	22	31	Pen	18	26	Wig	1	32
	Jug	6	18	Bib	2	22	Hem	4	17
	Pod	3	26	Col	7	30	Nip	3	28
	Bin	9	38	Ted	7	22	Tub	13	23
	Gum	14	21	Dig	10	22	Dot	13	30
	Lob	1	28	Hop	2	27	Yen	3	22
	Mean	8.0	26.1	Mean	8.7	26.0	Mean	7.9	26.3
Materials B	Din	1	38	Bag	42	31	Jam	6	26
	Ham	19	30	Gem	4	14	Log	11	25
	Mop	3	24	Mob	10	22	Pot	28	35
	Web	6	11	Pet	8	32	Fad	2	21
	Rig	5	26	Nil	1	31	Tin	12	39
	Net	34	34	Sod	3	28	Zip	1	19
	Cod	6	31	Dip	6	26	Bet	20	36
	Mean	10.6	27.7	Mean	10.6	26.3	Mean	11.4	28.7

Appendix 4(b): Nonword Sets used in Experiment 4

	Similar Nonword Sets					
	Filler 1		Filler 2		Hebb	
	Nonword	Neighbourhood Size	Nonword	Neighbourhood Size	Nonword	Neighbourhood Size
Materials A	Pef	12	Bem	16	Vem	7
	Pez	9	Bez	13	Vek	15
	Pem	12	Bep	13	Vep	6
	Sef	15	Nef	11	Lem	17
	Sem	16	Nep	14	Lep	15
	Seb	13	Nez	8	Lef	16
	Mean	12.8	Mean	12.5	Mean	12.7
Materials B	Gog	17	Yol	13	Jom	17
	Goz	8	Yof	6	Jol	18
	Gom	13	Yom	9	Jof	10
	Fom	13	Hoz	11	Zom	6
	Foz	11	Hom	18	Zof	3
	Fov	5	Hof	12	Zog	9
	Mean	11.2	Mean	11.5	Mean	10.5

Appendix 4(b): Nonword Sets used in Experiment 4 (continued)

	Distinct Nonword Sets					
	Filler 1		Filler 2		Hebb	
	Nonword	Neighbourhood Size	Nonword	Neighbourhood Size	Nonword	Neighbourhood Size
Materials A	Gid	17	Beb	14	Leb	12
	Vog	12	Jal	14	Taf	14
	Zup	6	Vig	13	Som	18
	Dal	17	Poz	12	Zid	9
	Mef	8	Fup	12	Vus	10
	Tez	8	Yem	8	Hez	13
	Mean	11.3	Mean	12.1	Mean	12.7
Materials B	Zep	6	Dif	18	Gup	13
	Sof	14	Mep	8	Kem	11
	Fub	19	Vom	7	Neg	12
	Vas	11	Jus	12	Fid	19
	Nid	17	Tal	19	Jav	10
	Heg	15	Zeg	6	Zos	8
	Mean	13.7	Mean	11.7	Mean	12.2

Appendix 4(c): Results of Statistical Analyses to Confirm Parallel Patterns of Results in Experiment 4

(i) Summary of ANOVAs conducted in order to confirm parallel patterns of results across materials A and B for each of F_1, F_2 and H_1 for word and nonword sequences.

Word Sequences

Filler 1: The main effect of materials and each of the two-way interactions failed to reach significance (all *p*s>.05). However, the phonological similarity x trials x materials interaction attained significance ($F(3,42) = 2.91$, $MSe = 0.655$, $p=.046$, $\eta_p^2=.17$). Further independent analyses based on each materials set demonstrated this three-way interaction was due to the emergence of a significant PSE for materials B only ($F(1,7) = 2.79$, $MSe = 4.087$, $p=.14$, for materials A; $F(1,7) = 9.80$, $MSe = 1.250$, $p=.017$, $r=.76$, for materials B). However, recall performance was higher for distinct than similar word sequences for both materials A and B. Indeed, the overall mean difference between these two types of word sequence was identical for each materials set (mean difference of 0.9), suggesting these contrasting results were simply due to larger degrees of variation in materials A. On this basis that the three-way interaction was only marginally significant and the small sample size (n=8), it was therefore decided to collapse across materials A and B.

Filler 2: The main effect of materials and all interactions failed to reach significance (all *p*s>.05). The data were therefore collapsed across materials A and B.

Hebb: The main effect of materials reached significance ($F(1,14) = 10.35$, $MSe = 3.543$, $p=.006$, $r=.65$), demonstrating better performance for materials B. All interactions failed to reach significance (all *p*s>.05). Given the lack of significant interactions, it remained appropriate to collapse over materials A and B.

Appendix 4(c): Results of Statistical Analyses to Confirm Parallel Patterns of Results in Experiment 4 (continued)

Nonword Sequences

Filler 1: The main effect of materials and all interactions failed to reach significance (all *ps*>.05). The data were therefore collapsed across materials A and B.

Filler 2: The main effect of materials failed to reach significance (p>.05). However, a significant trials x materials interaction emerged ($F(3,42) = 3.72$, $MSe = 0.590$, $p=.018$, $\eta_p^2=.21$). Simple main effects analysis revealed learning over trials for materials A ($F(3,42) = 3.56$, $MSe = 0.59$, $p=.022$) but not for materials B ($F(3,42) = 2.03$, $MSe = 0.59$, $p=.12$). However, significantly better performance was observed for materials A at Trial 4 ($F(1,14) = 4.81$, $MSe = 1.36$, $p=.046$), with equivalent performance at Trials 1 to 3 (all *ps*>.05). All remaining interactions failed to reach significance (all *ps*>.05). Given that the learning over trials observed in materials A was driven by performance at Trial 4 only, and the small sample size (n=8), the data were collapsed over materials A and B.

Hebb: The main effect of materials and all interactions failed to reach significance (all *ps*>.05). The data were therefore collapsed across materials A and B.

(i) Summary of ANOVAs conducted in order to confirm parallel patterns of results across F_1 and F_2 for word and nonword sequences.

Word Sequences: The main effect of filler list and all interactions failed to reach significance (all *ps*>.05). The data were therefore collapsed across F_1 and F_2.

Nonword Sequences: The main effect of filler list attained significance ($F(1,15) = 8.12$, $MSe = 1.018$, $p=.012$, $r=.59$), demonstrating better performance for F_2. All interactions failed to reach significance (all *ps*>.05). Given the lack of significant interactions, it remained appropriate to collapse over F_1 and F_2.

Appendix 4(d): Error Proportions for Word and Nonword Sequences in Experiment 4

Table 4.1(d): *Mean proportions (and standard deviations) for the three error types for Hebb and filler lists at each trial for similar and distinct word conditions*

		List Type/Error Type					
		Filler			Hebb		
Condition	Trials	Item	Order	Rep*	Item	Order	Rep*
Similar Words	1	.496 (.108)	.492 (.175)	.025 (.025)	.464 (.173)	.395 (.223)	.071 (.090)
	2	.469 (.133)	.471 (.241)	.047 (.062)	.397 (.163)	.375 (.202)	.076 (.112)
	3	.446 (.124)	.502 (.215)	.049 (.057)	.335 (.137)	.259 (.209)	.076 (.066)
	4	.404 (.126)	.421 (.179)	.060 (.061)	.322 (.238)	.321 (.304)	.049 (.057)
	Mean *(SD)*	.454 (.096)	.470 (.149)	.045 (0.38)	.379 (.131)	.338 (.156)	.068 (.062)
Distinct Words	1	.540 (.186)	.207 (.162)	.009 (.016)	.500 (.140)	.201 (.240)	.004 (.018)
	2	.440 (.165)	.226 (.121)	.022 (.029)	.357 (.181)	.159 (.174)	.009 (.024)
	3	.442 (.199)	.276 (.163)	.016 (.026)	.299 (.166)	.104 (.146)	.018 (.041)
	4	.375 (.188)	.259 (.181)	.007 (.015)	.196 (.194)	.052 (.097)	.009 (.024)
	Mean *(SD)*	.449 (.170)	.244 (.118)	.013 (.014)	.338 (.146)	.129 (.123)	.010 (.016)

* Repetition Errors

Appendix 4(d): Error Proportions for Word and Nonword Sequences in Experiment 4 (continued)

Table 4.2(e): *Mean proportions (and standard deviations) for item error sub-types for Hebb and filler lists at each trial for similar and distinct word conditions*

		List Type/Item Error Sub-Type							
		Filler				Hebb			
Condition	Trials	Total	Omissions	IEI*	EEI#	Total	Omissions	IEI*	EEI#
Similar Words	1	.496	.429	.022	.045	.464	.402	.027	.036
		(.108)	(.102)	(.043)	(.042)	(.173)	(.182)	(.036)	(.052)
	2	.469	.380	.056	.034	.397	.268	.080	.049
		(.133)	(.140)	(.078)	(.040)	(.163)	(.178)	(.119)	(.077)
	3	.446	.373	.054	.020	.335	.223	.076	.036
		(.124)	(.145)	(.060)	(.022)	(.137)	(.167)	(.115)	(.045)
	4	.404	.326	.051	.027	.322	.237	.045	.040
		(.126)	(.161)	(.055)	(.038)	(.238)	(.195)	(.078)	(.078)
	Mean	.454	.377	.046	.031	.379	.282	.057	.040
	(SD)	(.096)	(.118)	(.051)	(.026)	(.131)	(.141)	(.063)	(.041)
Distinct Words	1	.540	.398	.029	.114	.500	.380	.040	.080
		(.186)	(.154)	(.030)	(.112)	(.140)	(.137)	(.069)	(.063)
	2	.440	.308	.063	.069	.357	.259	.045	.054
		(.165)	(.171)	(.066)	(.066)	(.181)	(.179)	(.058)	(.071)
	3	.442	.326	.069	.047	.299	.183	.080	.036
		(.199)	(.169)	(.075)	(.053)	(.166)	(.138)	(.086)	(.045)
	4	.375	.299	.051	.025	.196	.152	.031	.013
		(.188)	(.161)	(.049)	(.043)	(.194)	(.158)	(.058)	(.029)
	Mean	.449	.333	.053	.064	.338	.243	.049	.046
	(SD)	(.170)	(.151)	(.044)	(.056)	(.146)	(.124)	(.052)	(.033)

* Intra-experimental intrusion errors

\# Extra-experimental intrusion errors

Appendix 4(d): Error Proportions for Word and Nonword Sequences in Experiment 4 (continued)

Table 4.3(d): *Mean proportions (and standard deviations) for the three error types for Hebb and filler lists at each trial for similar and distinct nonword conditions*

		List Type/Error Type					
		Filler			Hebb		
Condition	Trials	Item	Order$^+$	Rep*	Item	Order$^+$	Rep*
Similar Nonwords	1	.466 (.141)	.342 (.186)	.036 (.050)	.516 (.150)	.557 (.188)	.047 (.074)
	2	.461 (.108)	.264 (.172)	.050 (.038)	.542 (.197)	.480 (.394)	.042 (.053)
	3	.458 (.075)	.326 (.204)	.055 (.071)	.448 (.192)	.364 (.317)	.078 (.078)
	4	.474 (.076)	.392 (.230)	.042 (.048)	.479 (.203)	.499 (.305)	.057 (.066)
	Mean *(SD)*	.465 (.067)	.331 (.142)	.046 (.035)	.496 (.146)	.475 (.219)	.056 (.039)
Distinct Nonwords	1	.633 (.105)	.127 (.109)	.003 (.011)	.724 (.206)	.227 (.308)	.021 (.048)
	2	.685 (.133)	.078 (.116)	.008 (.017)	.620 (.161)	.152 (.229)	.010 (.028)
	3	.604 (.114)	.112 (.127)	.008 (.017)	.609 (.223)	.023 (.075)	.016 (.045)
	4	.612 (.141)	.119 (.148)	.005 (.014)	.536 (.294)	.011 (.038)	-
	Mean *(SD)*	.633 (.101)	.109 (.066)	.006 (.011)	.622 (.196)	.103 (.117)	.012 (.024)

* Repetition Errors

$^+$ Based on 11 participants.

Appendix 4(d): Error Proportions for Word and Nonword Sequences in Experiment 4 (continued)

Appendix 4.4(e): *Mean proportions (and standard deviations) for item error sub-types for Hebb and filler lists at each trial for similar and distinct nonword conditions*

		List Type/Item Error Sub-Type							
		Filler				Hebb			
Condition	Trials	Total	Omissions	IEI*	EEI#	Total	Omissions	IEI*	EEI#
Similar Nonwords	1	.466 (.141)	.318 (.131)	.029 (.050)	.120 (.087)	.516 (.150)	.359 (.166)	.026 (.066)	.130 (.086)
	2	.461 (.108)	.281 (.153)	.060 (.050)	.120 (.091)	.542 (.197)	.365 (.161)	.073 (.096)	.104 (.108)
	3	.458 (.075)	.287 (.120)	.047 (.059)	.125 (.070)	.448 (.192)	.240 (.155)	.068 (.076)	.141 (.113)
	4	.474 (.076)	.287 (.093)	.068 (.045)	.120 (.094)	.479 (.203)	.261 (.174)	.094 (.096)	.125 (.086)
	Mean *(SD)*	.465 (.067)	.293 (.095)	.051 (.037)	.121 (.065)	.496 (.146)	.306 (.129)	.065 (.059)	.125 (.055)
Distinct Nonwords	1	.633 (.105)	.242 (.128)	.029 (.029)	.362 (.166)	.724 (.206)	.245 (.138)	.057 (.073)	.422 (.210)
	2	.685 (.133)	.266 (.119)	.044 (.042)	.375 (.155)	.620 (.161)	.193 (.113)	.047 (.074)	.380 (.190)
	3	.604 (.114)	.268 (.115)	.037 (.040)	.300 (.165)	.609 (.223)	.172 (.157)	.052 (.074)	.385 (.223)
	4	.612 (.141)	.255 (.143)	.026 (.026)	.331 (.182)	.536 (.294)	.172 (.138)	.047 (.068)	.318 (.213)
	Mean *(SD)*	.633 (.101)	.258 (.104)	.034 (.019)	.342 (.142)	.622 (.196)	.195 (.095)	.051 (.047)	.376 (.183)

* Intra-experimental intrusion errors
\# Extra-experimental intrusion errors

Appendix 5(a): Error Proportions for Word and Nonword Sequences in Experiment 5

Table 5.1(a): *Mean proportions (and standard deviations) for the three error types for Hebb and filler lists at each trial for similar and distinct word conditions*

		List Type/Error Type					
		Filler			Hebb		
Condition	Trials	Item	Order	Rep*	Item	Order	Rep*
Similar Words	1	.443	.490	.042	.441	.515	.022
		(.143)	(.186)	(.048)	(.163)	(.291)	(.033)
	2	.407	.574	.036	.376	.564	.040
		(.130)	(.210)	(.044)	(.179)	(.291)	(.056)
	3	.388	.498	.033	.382	.525	.034
		(.119)	(.203)	(.032)	(.180)	(.247)	(.047)
	4	.415	.475	.039	.370	.440	.028
		(.150)	(.234)	(.041)	(.187)	(.264)	(.047)
	Mean	.413	.509	.037	.392	.511	.031
	(SD)	(.123)	(.162)	(.027)	(.158)	(.210)	(.028)
Distinct Words	1	.436	.274	.023	.351	.305	.019
		(.156)	(.181)	(.028)	(.124)	(.294)	(.039)
	2	.320	.391	.029	.248	.302	.037
		(.150)	(.240)	(.033)	(.176)	(.238)	(.068)
	3	.320	.363	.019	.248	.250	.028
		(.139)	(.201)	(.030)	(.188)	(.281)	(.042)
	4	.280	.326	.036	.255	.238	.028
		(.139)	(.200)	(.043)	(.197)	(.262)	(.047)
	Mean	.339	.338	.027	.276	.274	.028
	(SD)	(.127)	(.161)	(.020)	(.135)	(.238)	(.039)

* Repetition Errors

Appendix 5(a): Error Proportions for Word and Nonword Sequences in Experiment 5 (continued)

Table 5.2(a): *Mean proportions (and standard deviations) for item error sub-types for Hebb and filler lists at each trial for similar and distinct word conditions*

		List Type/Item Error Sub-Type							
		Filler				Hebb			
Condition	Trials	Total	Omissions	IEI*	EEI#	Total	Omissions	IEI*	EEI#
Similar Nonwords	1	.443 (.143)	.410 (.145)	-	.033 (.047)	.441 (.163)	.413 (.165)	-	.028 (.052)
	2	.407 (.130)	.385 (128)	-	.022 (.037)	.376 (.179)	.348 (.175)	-	.028 (.047)
	3	.388 (.119)	.371 (115)	-	.017 (.032)	.382 (.180)	.366 (.174)	-	.015 (.037)
	4	.415 (.150)	.383 (.146)	-	.031 (.045)	.370 (.187)	.345 (.166)	-	.025 (.059)
	Mean *(SD)*	.413 (.123)	.387 (.121)	-	.026 (.034)	.392 (.158)	.368 (.151)	-	.024 (.036)
Distinct Nonwords	1	.436 (.156)	.362 (.126)	-	.075 (.076)	.351 (.124)	.326 (.112)	-	.025 (.041)
	2	.320 (.150)	.294 (.140)	-	.026 (.046)	.248 (.176)	.220 (.174)	-	.028 (.056)
	3	.320 (.139)	.295 (.133)	-	.025 (.041)	.248 (.188)	.230 (.168)	-	.019 (.039)
	4	.280 (.139)	.258 (.123)	-	.022 (.038)	.255 (.197)	.227 (.189)	-	.028 (.056)
	Mean *(SD)*	.339 (.127)	.302 (.112)	-	.037 (.042)	.276 (.135)	.251 (.125)	-	.025 (.035)

* Intra-experimental intrusion errors
# Extra-experimental intrusion errors

Appendix 5(a): Error Proportions for Word and Nonword Sequences in Experiment 5 (continued)

Table 5.3(a): *Mean proportions (and standard deviations) for the three error types for Hebb and filler lists at each trial for similar and distinct nonword conditions*

		List Type/Error Type					
		Filler			Hebb		
Condition	Trials	Item	Order⁺	Rep*	Item	Order⁺	Rep*
Similar Nonwords	1	.464 (.103)	.331 (.152)	.035 (.041)	.467 (.109)	.534 (.232)	.043 (.061)
	2	.453 (.159)	.456 (.269)	.053 (.049)	.427 (.141)	.522 (.274)	.043 (.070)
	3	.426 (.104)	.413 (.156)	.044 (.053)	.467 (.170)	.452 (.237)	.058 (.085)
	4	.420 (.143)	.395 (.175)	.053 (.052)	.464 (.163)	.367 (.310)	.062 (.068)
	Mean *(SD)*	.441 (.098)	.399 (.107)	.046 (.033)	.457 (.091)	.469 (.159)	.052 (.047)
Distinct Nonwords	1	.616 (.133)	.141 (.113)	.005 (.014)	.591 (.188)	.196 (.212)	.014 (.041)
	2	.505 (.173)	.180 (.183)	.013 (.029)	.446 (.189)	.233 (.260)	.014 (.041)
	3	.518 (.167)	.207 (.207)	.011 (.019)	.420 (.218)	.183 (.179)	.014 (.041)
	4	.447 (.188)	.184 (.170)	.013 (.020)	.373 (.226)	.210 (.231)	.022 (.045)
	Mean *(SD)*	.522 (.131)	.178 (.116)	.010 (.011)	.457 (.169)	.206 (.149)	.016 (.024)

* Repetition Errors

⁺ Based on 21 participants.

Appendix 5(a): Error Proportions for Word and Nonword Sequences in Experiment 5 (continued)

Table 5.4(a): *Mean proportions (and standard deviations) for **item error** sub-types for Hebb and filler lists at each trial for similar and distinct **nonword conditions***

		List Type/Item Error Sub-Type							
		Filler				Hebb			
Condition	Trials	Total	Omissions	IEI*	EEI#	Total	Omissions	IEI*	EEI#
Similar	1	.464	.384	-	.080	.467	.362	.004	.101
Nonwords		(.103)	(.112)		(.084)	(.109)	(.139)	(.017)	(.104)
	2	.453	.357	.004	.092	.427	.326	.004	.098
		(.159)	(.135)	(.012)	(.088)	(.141)	(.163)	(.017)	(.120)
	3	.426	.342	.007	.076	.467	.341	.004	.123
		(.104)	(.110)	(.020)	(.096)	(.170)	(.168)	(.017)	(.128)
	4	.420	.322	.002	.096	.464	.348	-	.116
		(.143)	(.140)	(.009)	(.108)	(.163)	(.114)		(.135)
	Mean	.441	.351	.003	.086	.457	.344	.003	.110
	(SD)	(.098)	(.098)	(.010)	(.079)	(.091)	(.119)	(.010)	(.100)
Distinct	1	.616	.312	.004	.301	.591	.265	.004	.322
Nonwords		(.133)	(.161)	(.012)	(.138)	(.188)	(.154)	(.017)	(.169)
	2	.505	.286	.002	.217	.446	.221	-	.225
		(.173)	(.126)	(.009)	(.149)	(.189)	(.119)		(.168)
	3	.518	.304	-	.214	.420	.203	.007	.210
		(.167)	(.117)		(.151)	(.218)	(.139)	(.035)	(.210)
	4	.447	.241	.002	.205	.373	.152	.007	.214
		(.188)	(.149)	(.009)	(.164)	(.226)	(.134)	(.024)	(.193)
	Mean	.522	.286	.002	.234	.457	.210	.005	.243
	(SD)	(.131)	(.109)	(.007)	(.125)	(.169)	(.094)	(.018)	(.156)

* Intra-experimental intrusion errors

\# Extra-experimental intrusion errors

Appendix 6(a): Nonword Sets used in Experiment 6

	Similar Nonword Sets					
	List 1		List 2		List 3	
	Nonword	Neighbourhood Size	Nonword	Neighbourhood Size	Nonword	Neighbourhood Size
Materials A	Pef	12	Bem	16	Vem	7
	Pez	9	Bez	13	Vek	15
	Pem	12	Bep	13	Vep	6
	Sef	15	Nep	14	Lep	15
	Sem	16	Nez	8	Lef	16
	Mean	12.8	Mean	12.8	Mean	11.8
Materials B	Gog	17	Yol	13	Jom	17
	Goz	8	Yof	6	Jol	18
	Gom	13	Yom	9	Jof	10
	Fom	13	Hom	18	Zom	6
	Foz	11	Hof	12	Zog	9
	Mean	12.4	Mean	11.6	Mean	12.0

Appendix 6(a): Nonword Sets used in Experiment 6 (continued)

	Distinct Nonword Sets					
	List 1		List 2		List 3	
	Nonword	Neighbourhood Size	Nonword	Neighbourhood Size	Nonword	Neighbourhood Size
Materials A	Gid	17	Jal	14	Leb	12
	Vog	12	Vig	13	Taf	14
	Zup	6	Poz	12	Som	18
	Dal	17	Fup	12	Zid	9
	Tez	8	Yem	8	Vus	10
	Mean	12.0	Mean	11.8	Mean	12.6
Materials B	Zep	6	Dif	18	Gup	13
	Sof	14	Mep	8	Neg	12
	Fub	19	Vom	7	Fid	19
	Vas	11	Jus	12	Jav	10
	Nid	17	Tal	19	Zos	8
	Mean	13.4	Mean	12.8	Mean	12.4

Appendix 6(b): Error Proportions for Word and Nonword Sequences in Experiment 6

Table 6.1(b): *Mean proportions (and standard deviations) for the three error types for Hebb and filler lists at each trial for similar and distinct word conditions*

		List Type/Error Type					
		Filler			Hebb		
Condition	Trials	Item	Order$^+$	Rep*	Item	Order	Rep*
Similar Words	1	.509	.427	.023	.547	.356	.025
		(.095)	(.207)	(.028)	(.193)	(.248)	(.051)
	2	.509	.393	.017	.450	.319	.025
		(.103)	(.181)	(.024)	(.186)	(.253)	(.041)
	3	.526	.442	.020	.370	.299	.040
		(.101)	(.224)	(.030)	(.217)	(.250)	(.052)
	4	.547	.408	.025	.311	.237	.031
		(.109)	(.166)	(.029)	(.227)	(.231)	(.047)
	Mean	.523	.417	.021	.419	.303	.030
	(SD)	(.084)	(.144)	(.019)	(.185)	(.190)	(.030)
Distinct Words	1	.505	.203	.006	.556	.130	.019
		(.129)	(.169)	(.018)	(.167)	(.154)	(.039)
	2	.489	.191	.014	.370	.114	.015
		(.112)	(.164)	(.021)	(.199)	(.164)	(.037)
	3	.480	.231	.011	.264	.082	.025
		(.150)	(.171)	(.020)	(.187)	(.126)	(.051)
	4	.463	.175	.012	.267	.058	.006
		(.116)	(.151)	(.020)	(.223)	(.102)	(.020)
	Mean	.484	.200	.011	.364	.096	.016
	(SD)	(.103)	(.117)	(.012)	(.175)	(.102)	(.030)

$^+$ Based on 22 participants.

* Repetition Errors

Appendix 6(b): Error Proportions for Word and Nonword Sequences in Experiment 6 (continued)

Table 6.2(b): *Mean proportions (and standard deviations) for item error sub-types for Hebb and filler lists at each trial for similar and distinct word conditions*

		List Type/Item Error Sub-Type							
		Filler				Hebb			
Condition	Trials	Total	Omissions	IEI*	EEI#	Total	Omissions	IEI*	EEI#
Similar Nonwords	1	.509	.461	.012	.036	.547	.432	.056	.059
		(.095)	(.096)	(.026)	(.044)	(.193)	(.130)	(.080)	(.077)
	2	.509	.429	.056	.025	.450	.366	.034	.050
		(.103)	(.122)	(.049)	(.033)	(.186)	(.148)	(.042)	(.066)
	3	.526	.439	.047	.040	.370	.273	.062	.034
		(.101)	(.140)	(.053)	(.051)	(.217)	(.158)	(.069)	(.064)
	4	.547	.457	.053	.037	.311	.217	.056	.037
		(.109)	(.125)	(.056)	(.045)	(.227)	(.180)	(.071)	(.060)
	Mean	.523	.446	.042	.035	.419	.322	.052	.045
	(SD)	(.084)	(.104)	(.035)	(.029)	(.185)	(.126)	(.052)	(.053)
Distinct Nonwords	1	.505	.370	.026	.109	.556	.385	.040	.130
		(.129)	(.126)	(.033)	(.095)	(.167)	(.132)	(.071)	(.117)
	2	.489	.346	.056	.087	.370	.264	.031	.074
		(.112)	(.129)	(.047)	(.098)	(.199)	(.151)	(.068)	(.102)
	3	.480	.340	.068	.071	.264	.174	.046	.043
		(.150)	(.116)	(.054)	(.073)	(.187)	(.147)	(.076)	(.067)
	4	.463	.340	.054	.068	.267	.180	.028	.059
		(.116)	(.121)	(.060)	(.065)	(.223)	(.185)	(.052)	(.093)
	Mean	.484	.349	.051	.084	.364	.251	.036	.077
	(SD)	(.103)	(.099)	(.036)	(.070)	(.175)	(.127)	(.050)	(.081)

* Intra-experimental intrusions

\# Extra-experimental intrusions

Appendix 6(b): Error Proportions for Word and Nonword Sequences in Experiment 6 (continued)

Table 6.3(b): *Mean proportions (and standard deviations) for the three error types for Hebb and filler lists at each trial for similar and distinct nonword conditions*

		List Type/Error Type					
		Filler			Hebb		
Condition	Trials	Item	Order^{+}	Rep*	Item	Order^{+}	Rep*
Similar Nonwords	1	.344	.202	.044	.429	.263	.050
		(.121)	(.192)	(.063)	(.216)	(.217)	(.078)
	2	.425	.324	.044	.408	.221	.058
		(.149)	(.181)	(.043)	(.212)	(.282)	(.088)
	3	.450	.229	.025	.392	.232	.071
		(.157)	(.163)	(.039)	(.232)	(.277)	(.086)
	4	.463	.286	.021	.408	.123	.054
		(.152)	(.192)	(.033)	(.267)	(.153)	(.072)
	Mean	.420	.260	.033	.409	.210	.058
	(SD)	(.117)	(.122)	(.025)	(.193)	(.158)	(.064)
Distinct Nonwords	1	.610	.136	.004	.550	.089	-
		(.201)	(.196)	(.014)	(.269)	(.198)	
	2	.581	.083	.002	.421	.047	.004
		(.196)	(.169)	(.010)	(.215)	(.105)	(.020)
	3	.513	.087	.006	.379	.023	.004
		(.200)	(.144)	(.017)	(.208)	(.063)	(.020)
	4	.506	.015	.002	.383	.026	-
		(.227)	(.044)	(.010)	(.237)	(.073)	
	Mean	.553	.081	.004	.433	.046	.002
	(SD)	(.180)	(.106)	(.007)	(.205)	(.089)	(.007)

$^{+}$ Based on 16 participants

* Repetition Errors

Appendix 6(b): Error Proportions for Word and Nonword Sequences in Experiment 6 (continued)

Table 6.4(b): *Mean proportions (and standard deviations) for item error sub-types for Hebb and filler lists at each trial for similar and distinct* ***nonword*** *conditions*

		List Type/Item Error Sub-Type							
		Filler				Hebb			
Condition	Trials	Total	Omissions	IEI*	EEI#	Total	Omissions	IEI*	EEI#
Similar	1	.344	.208	.038	.098	.429	.217	.029	.183
Nonwords		(.121)	(.101)	(.073)	(.088)	(.216)	(.143)	(.055)	(.204)
	2	.425	.219	.040	.167	.408	.163	.067	.179
		(.149)	(.096)	(.055)	(.124)	(.212)	(.166)	(.092)	(.169)
	3	.450	.190	.058	.202	.392	.113	.050	.229
		(.157)	(.111)	(.060)	(.157)	(.232)	(.115)	(.083)	(.205)
	4	.463	.194	.073	.196	.408	.113	.058	.238
		(.152)	(.142)	(.066)	(.126)	(.267)	(.115)	(.093)	(.236)
	Mean	.420	.203	.052	.166	.409	.151	.051	.207
	(SD)	(.117)	(.085)	(.046)	(.107)	(.193)	(.105)	(.067)	(.169)
Distinct	1	.610	.108	.023	.479	.550	.154	.021	.375
Nonwords		(.201)	(.101)	(.033)	(.200)	(.269)	(.161)	(.051)	(.231)
	2	.581	.121	.017	.444	.421	.075	.025	.321
		(.196)	(.110)	(.032)	(.195)	(.215)	(.111)	(.053)	(.182)
	3	.513	.115	.035	.363	.379	.029	.025	.325
		(.200)	(.099)	(.038)	(.178)	(.208)	(.062)	(.044)	(.209)
	4	.506	.096	.038	.373	.383	.025	.033	.325
		(.227)	(.088)	(.063)	(.181)	(.237)	(.068)	(.064)	(.233)
	Mean	.553	.110	.028	.415	.433	.071	.026	.336
	(SD)	(.180)	(.081)	(.027)	(.161)	(.205)	(.078)	(.030)	(.178)

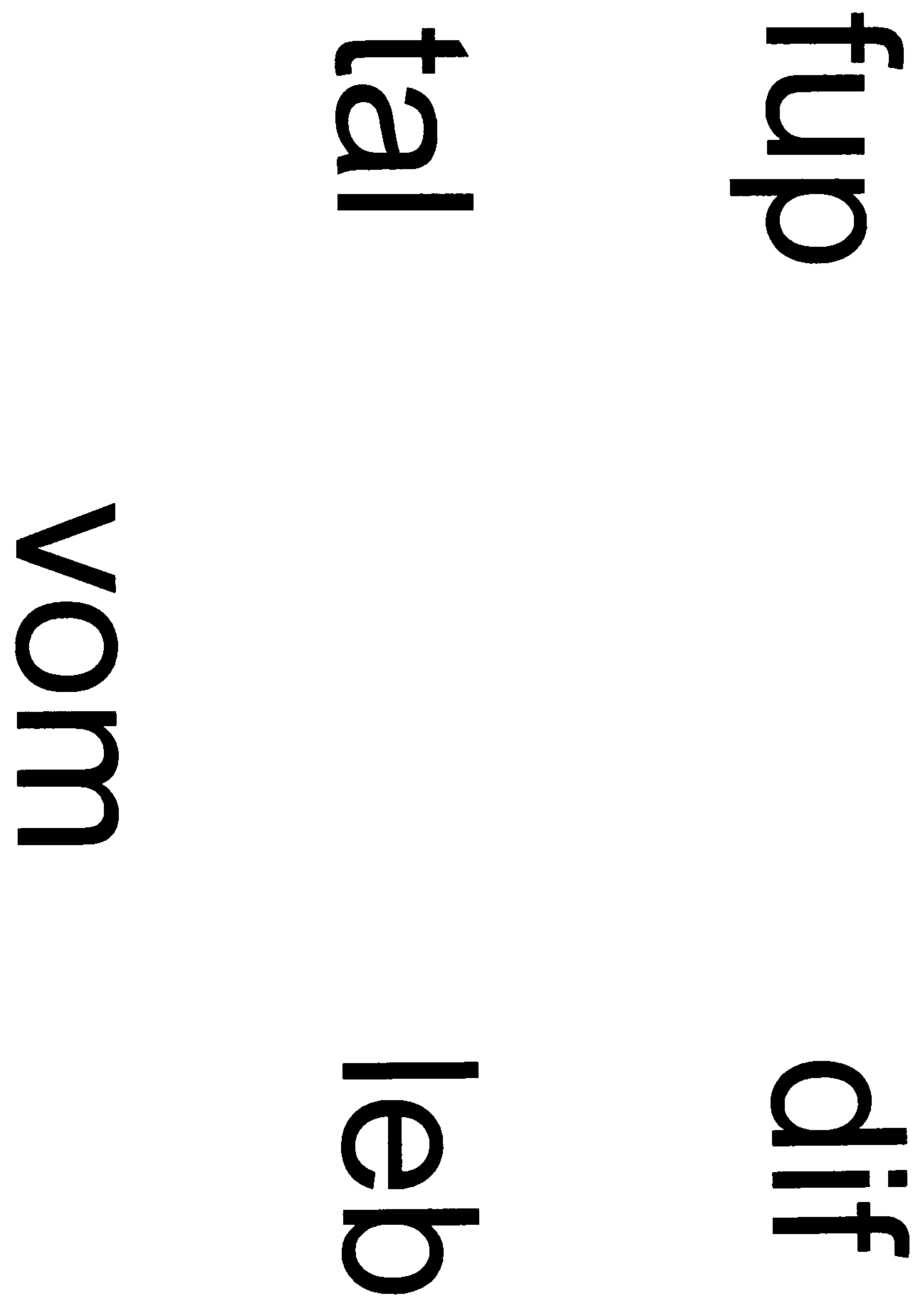
* Intra-experimental intrusions

\# Extra-experimental intrusions

Appendix 7(a): Nonword Sets used in Experiment 7

	List 1		List 2		List 3	
	Nonword	Neighbourhood Size	Nonword	Neighbourhood Size	Nonword	Neighbourhood Size
Similar Nonword Sets	Pef	12	Bem	16	Vem	7
	Pez	9	Bez	13	Vek	15
	Pem	12	Bep	13	Vep	6
	Sef	15	Nef	11	Lem	17
	Sem	16	Nep	14	Lep	15
	Seb	13	Nez	8	Lef	16
	Mean	12.8	Mean	12.5	Mean	12.7
Distinct Nonword Sets	Zep	6	Dif	18	Gup	13
	Poz	12	Leb	12	Kem	11
	Fub	19	Vom	7	Beb	14
	Jav	10	Fup	12	Vig	13
	Nid	17	Tal	19	Dal	17
	Heg	15	Zeg	6	Zos	8
	Mean	13.2	Mean	12.3	Mean	12.7

Appendix 7(b): Example of Visual Array used in Experiment 7



Appendix 7(c): Error Proportions for Nonword Sequences in Experiment 7

Table 7.1(c): *Mean proportions (and standard deviations) for the three error types for Hebb and filler lists at each trial for similar and distinct nonword conditions*

		List Type/Error Type					
		Filler			Hebb		
Condition	Trials	Item	Order	Rep*	Item	Order	Rep*
Similar Nonwords	1	.030 (.047)	.502 (.153)	.021 (.041)	.031 (.054)	.529 (.232)	.024 (.058)
	2	.021 (.027)	.487 (.171)	.024 (.030)	.031 (.073)	.574 (.186)	.021 (.044)
	3	.037 (.078)	.516 (.141)	.024 (.035)	.021 (.086)	.525 (.226)	.049 (.073)
	4	.033 (.061)	.509 (.184)	.024 (.044)	.028 (.103)	.467 (.304)	.014 (.040)
	Mean *(SD)*	.030 (.041)	.503 (.116)	.023 (.028)	.028 (.072)	.524 (.193)	.027 (.033)
Distinct Nonwords	1	.043 (.063)	.326 (.160)	.005 (.019)	.049 (.069)	.292 (.220)	-
	2	.040 (.050)	.315 (.149)	.004 (.012)	.042 (.085)	.227 (.220)	.007 (.023)
	3	.050 (.078)	.233 (.131)	.004 (.012)	.038 (.107)	.179 (.187)	-
	4	.038 (.065)	.239 (.169)	.004 (.012)	.062 (.131)	.159 (.198)	.003 (.017)
	Mean *(SD)*	.043 (.054)	.279 (.111)	.004 (.011)	.048 (.067)	.214 (.166)	.003 (.007)

* Repetition Errors

References

Addis, K. M. & Kahana, M. J. (2004). Decomposing serial learning: What is missing from the learning curve? *Psychonomic Bulletin and Review, 11,* 118-124.

Allen, R. & Hulme, C. (2006). Speech and language processing mechanisms in verbal serial recall. *Journal of Memory and Language, 55,* 64-88.

Archibald, L. M. D. & Gathercole, S. E. (in press). Nonword repetition and serial recall: Equivalent measures of verbal short-term memory? *Applied Psycholinguistics.*

Archibald, L. M. D. & Gathercole, S. E. (2006). Short-term and working memory in specific language impairment. *International Journal of Communication Disorders, 41,* 675-693.

Atkins, P. W. B. & Baddeley, A. D. (1998). Working memory and distributed vocabulary learning. *Applied Psycholinguistics, 19,* 537-552.

Avons, S. E. (1998). Serial report and item recognition of novel visual patterns. *British Journal of Psychology, 89,* 285-308.

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

Atkinson, R. C. & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In. K. W. Spence (Ed.) *The psychology of learning and motivation: advances in research and theory.* (Vol. 2, pp. 89-195). New York: Academic Press.

Baayen, R. H., Piepenbrock, R. & van Rijn, H. (1993). *The CELEX Lexical Database* [CD-ROM]. Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania.

Baddeley, A. D. (1966a). Short-term memory for word sequences as a function of acoustic, semantic and formal similarity. *Quarterly Journal of Experimental Psychology, 18,* 362-365.

Baddeley, A. D. (1966b). The influence of acoustic and semantic similarity on long-term memory for word sequences. *Quarterly Journal of Experimental Psychology, 18*, 302-309.

Baddeley, A. D. (1986). *Working memory.* Oxford, England: Oxford University Press.

Baddeley, A. D. (1992). Working memory. *Science, 255,* 556-559.

Baddeley, A. D. (1993). Short-term phonological memory and long-term learning: A single case study. *European Journal of Cognitive Psychology, 5,* 129-148.

Baddeley, A. D. (2000a). The episodic buffer: a new component of working memory? *Trends in Cognitive Sciences, 4,* 417-423.

Baddeley, A. D. (2000b). The phonological loop and the irrelevant speech effect: Some comments on Neath (2000). *Psychonomic Bulletin and Review, 7,* 544-549.

Baddeley, A. D (2003). Working memory: Looking back and looking forward. *Nature Reviews Neuroscience, 4,* 829-839.

Baddeley, A. D. (2007). *Working memory, thought, and action.* Oxford, England: Oxford University Press.

Baddeley, A. D. & Andrade, J. (1994). Reversing the word length effect: A comment on Caplan, Rochon and Waters. *Quarterly Journal of Experimental Psychology, 47A,* 1047-1054.

Baddeley, A. D., Chincotta, D., Stafford, L. & Turk, D. (2002). Is the word length effect in STM entirely attributable to output delay? Evidence from serial recognition. *Quarterly Journal of Experimental Psychology, 55A,* 353-369.

Baddeley, A. D. & Dale, H. C. A. (1966). The effect of semantic similarity and retroactive interference in long- and short-term memory. *Journal of Verbal Learning and Verbal Behaviour, 5,* 417-420.

Baddeley, A. D., Gathercole, S. & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review, 105,* 158-173.

Baddeley, A. D. & Hitch, G. (1974). Working memory. In G. A. Bower (Ed.) *Recent advances in learning and motivation.* (Vol. 8, pp. 47-90). New York: Academic Press.

Baddeley, A. D. & Larsen, J. D. (2007a). The phonological loop unmasked? A comment on the evidence for a "perceptual-gestural" alternative. *Quarterly Journal of Experimental Psychology, 60,* 497-504.

Baddeley, A. D. & Larsen, J. D. (2007b). The phonological loop: Some answers and some questions. *Quarterly Journal of Experimental Psychology, 60,* 512-518.

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

Baddeley, A. D., Papagno, C. & Vallar, G. (1988). When long-term learning depends on short-term storage. *Journal of Memory and Language*, 27, 586-595.

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

Baddeley, A. D. & Warrington, E. K. (1970). Amnesia and the distinction between long- and short-term memory. *Journal of Verbal Learning and Verbal Behaviour, 9,* 176-189.

Baddeley, A. D. & Wilson, B. A. (1985). Phonological coding and short-term memory in patients without speech. *Journal of Memory and Language, 24,* 490-502.

Baddeley, A. D. & Wilson, B. A. (1994). When implicit learning fails: Amnesia and the problem of error elimination. *Neuropsychologia, 32,* 53-68.

Bartz, W. H. (1969). Repetition and the memory stores. *Journal of Experimental Psychology, 80,* 33-38.

Basso, A., Spinnler, H., Vallar, G. & Zanobio, M. E. (1982). Left hemisphere damage and selective impairment of auditory verbal short-term memory. A case study. *Neuropsychologia, 20,* 263-274.

Bishop, D. V. M. (1992). The underlying nature of specific language impairment. *Journal of Child Psychology and Child Psychiatry, 33,* 1-64.

Bishop, D. V. M., Adams, C. V. & Norbury, C. F. (2006). Distinct genetic influences on grammar and phonological short-term memory deficits: Evidence from 6-year-old twins. *Genes, Brain and Behaviour, 5,* 158-169.

Bishop, D. V. M., North, T. & Donlan, C. (1996). Nonword repetition as a behavioural maker of inherited language impairment: Evidence from a twin study. *Journal of Child Psychology and Psychiatry, 37,* 391-403.

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

Botting, N. & Conti-Ramsden, G. (2001). Non-word repetition and language development in children with specific language impairment (SLI). *International Journal of Language and Communication Disorders, 36,* 421-432.

Botvinick, M. M. & Huffstettler, S. (2006). *Sequence learning in short-term memory: Computational and empirical investigations of the Hebb effect.* Manuscript submitted for publication.

Botvinick, M. M. & Plaut, D. C. (2006). Short-term memory for serial order: A recurrent neural network model. *Psychological Review, 113,* 201-233.

Bower, G. H. (1961). Applications of a model to paired-associate learning. *Psychometrika, 26,* 255-280.

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

Bowey, J. A. (1996). On the association between phonological memory and receptive vocabulary in five-year-olds. *Journal of Experimental Child Psychology, 63,* 44-78.

Bowey, J. A. (1997). What does nonword repetition measure? A reply to Gathercole and Baddeley. *Journal of Experimental Child Psychology, 67,* 295-301.

Bowey, J. A. (2001). Nonword repetition and young children's receptive vocabulary: A longitudinal study. *Applied Psycholinguistics, 22,* 441-469.

Bowey, J. A. (2006). Clarifying the phonological processing account of nonword repetition. *Applied Psycholinguistics, 27,* 548-552.

Brown, R. A. (1973). *A first language: The early stages.* Cambridge, MA: Harvard University Press.

Brown, G. D. & Hulme, C. (1995). Modelling item length effects in memory span: No rehearsal needed? *Journal of Memory and Language, 34,* 594-624.

Brown, G. D., Preece, T. & Hulme, C. (2000). Oscillator-based memory for serial order. *Psychological Review, 107,* 127-181.

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

Burgess, N. & Hitch, G. J. (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. (2005). Computational models of working memory: Putting long-term memory into context. *Trends in Cognitive Sciences, 9,* 535-541.

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.

Caird, W. K. (1964). Reverberatory activity and memory disorder. *Nature, 201,* 1150.

Caplan, D., Rochon, E. & Waters, G. S. (1992). Articulatory and phonological determinants of word length effects in span tasks. *Quarterly Journal of Experimental Psychology, 45A,* 177-192.

Caplan, D. & Waters, G. S. (1994). Articulatory length and phonological similarity in span tasks: A reply to Baddeley and Andrade. *Quarterly Journal of Experimental Psychology, 47A,* 1055-1062.

Carey, S. (1978). The child as a word learner. In M. Halle, J. Bresnan & G. Miller (Eds.) *Linguistic theory and psychological reality.* Cambridge, MA: MIT Press.

Carlesimo, G. A., Galloni, F., Bonanni, R. & Sabbadini, M. (2006). Verbal short-term memory in individuals with congenital articulatory disorders: New empirical data and review of the literature. *Journal of Intellectual Disability Research, 50,* 81-91.

Cheung, H. (1996). Nonword span as a unique predictor of second-language vocabulary learning. *Developmental Psychology, 32,* 867-873.

Chiat, S. (2001). Mapping theories of developmental language impairment: Premises, predictions and evidence. *Language and Cognitive Processes, 16,* 113-142.

Chiat, S. (2006). The developmental trajectory of nonword repetition. *Applied Psycholinguistics, 27,* 552-556.

Cohen, R. L. & Johansson, B. S. (1967a). The activity trace in immediate memory: A re-evaluation. *Journal of Verbal Learning and Verbal Behaviour, 6,* 139-143.

Cohen, R. L. & Johansson, B. S. (1967b). Some relevant factors in the transfer of material from short-term to long-term memory. *Quarterly Journal of Experimental Psychology, 19,* 300-308.

Colle, H. A. & Welsh, A. (1976). Acoustic masking in primary memory. *Journal of Verbal Learning and Verbal Behaviour, 15,* 17-32.

Coltheart, M. (1981). The MRC Psycholinguistic Database. *Quarterly Journal of Experimental Psychology, 33A,* 497-505.

Conrad, R. (1964). Acoustic confusion and immediate memory. *British Journal of Psychology, 55,* 75-84.

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

Conrad, R. (1970). Short-term memory processes in the deaf. *British Journal of Psychology, 5,* 398-405.

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

Corsi, P. M. (1972). *Human memory and the medial temporal region of the brain.* Unpublished Thesis. McGill University, Montreal.

Couture, M. & Tremblay, S. (in press). Exploring the characteristics of the visuo-spatial Hebb repetition effect. *Memory and Cognition.*

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

Cowan, N., Nugent, L. D., Elliot, E. M. & Geer, T. (2000). Is there a temporal basis of the word length effect? A response to Service (1998). *Quarterly Journal of Experimental Psychology, 53A,* 647-660.

Cowan, N., Wood, N. L., Nugent, L. D. & Treisman, M. (1997). There are two word length effects in verbal short-term memory: Opposed effects of duration and complexity. *Psychological Science, 8,* 290-295.

Craik, F. I. M. & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behaviour, 11,* 671-684

Crano, W. D. & Mellon, P. M. (1978). Causal influence of teachers' expectations of children's academic performance: A cross-lagged panel analysis. *Journal of Educational Psychology, 70,* 39-49.

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

Cumming, N., Page, M. P. A., Norris, D. G. (2003). Testing a positional model of the Hebb effect. *Memory, 11,* 43-63.

Cumming, N., Page, M. P. A., Norris, D. G., McNeil, A. M. & Hitch, G. J. (2006). *Repetition spacing and order competition effects in the Hebb repetition task.* Unpublished manuscript.

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

Dale, H. C. A. & Baddeley, A. D. (1969). Acoustic similarity in long-term paired-associate learning. *Psychonomic Science, 16,* 209-211.

de Jong, P. F., Seveke, M-J. & van Veen, M. (2000). Phonological sensitivity and the acquisition of new words in children. *Journal of Experimental Child Psychology, 76,* 275-301.

Della Sala, S., Gray, C., Baddeley, A., Allamano, N. & Wilson, L. (1999). Pattern span: a tool for unwelding visuo-spatial memory. *Neuropsychologia, 37,* 1189-1199.

Della Sala, S., Gray, C., Baddeley, A. & Wilson, L. (1997). *The Visual Patterns Test: A new test of short-term visual recall.* Feltham, Suffolk: Thames Valley Test Company.

Della Sala, S. & Logie, R. H. (2002). Neuropsychological impairments of visual and spatial working memory. In A. D. Baddeley, M. D. Kopelman & B. A Wilson (Eds) *The Handbook of Memory Disorders* (2nd Edition) (pp. 271-292). West Sussex, England: John Wiley & Sons Ltd.

Destrebecqz, A. & Cleeremans, A. (2002). Temporal effects in sequence learning. In L. Jiménez (Ed.) *Attention and implicit learning* (pp. 181-214). Amsterdam: John Benjamins.

Dollaghan, C. A. (1987). Fast mapping in normal and language-impaired children. *Journal of Speech and Hearing Disorders, 52,* 218-222.

Dollaghan, C. A. & Campbell, T. F. (1998). Nonword repetition and child language impairment. *Journal of Speech, Language and Hearing Research, 41,* 1136-1146.

Drachman, D. A. & Arbit, J. (1966). Memory and the hippocampal complex II. Is memory a multiple process? *Archives of Neurology, 15,* 52-61.

Dunn, L. M. & Dunn, L. M. (1982). *The British Picture Vocabulary Scale*. Windsor. England: NFER-Nelson.

Ellis, N. & Beaton, A. (1993). Factors affecting the learning of foreign language vocabulary: Imagery keyword mediators and phonological short-term memory. *Quarterly Journal of Experimental Psychology, 46A,* 533-558.

Ellis Weismer, S. & Edwards, J. (2006). The role of phonological storage deficits in specific language impairment: A reconsideration. *Applied Psycholinguistics, 27,* 556-562.

Ellis Weismer, S. & Hesketh, L. (1996). Lexical learning by children with specific language impairment: Effects of linguistic input presented at varying speaking rates. *Journal of Speech and Hearing Research, 39,* 177-190.

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

Fallon, A. B., Dommett, K. M. & Tehan, G. (2005). *The role of covert rehearsal and recall in the production of the Hebb effect.* Unpublished manuscript.

Fallon, A. B., Groves, K. & Tehan, G. (1999). Phonological similarity and trace degradation in the serial recall task: When CAT helps RAT, but not MAN. *International Journal of Psychology, 34,* 301-307.

Farah, M. J. (1988). Is visual imagery really visual? Overlooked evidence from neuropsychology. *Psychological Review, 95,* 307-317.

Farah, M. J., Levine, D. N. & Calvanio, R. (1988). A case study of mental imagery deficit. *Brain and Cognition, 7/8,* 147-164.

Field, A. (2005). *Discovering statistics using SPSS* (2nd Edition). London: SAGE Publications Ltd.

Flavell, J. H., Beach, D. R. & Chinsky, J. M. (1966). Spontaneous verbal rehearsal in a memory task as a function of age. *Child Development, 37,* 283-299.

Gagnon, S., Bédard, M-J. & Turcotte, J. (2005). The effect of old age on supra-span learning of visuo-spatial sequences under incidental and intentional encoding instructions. *Brain and Cognition, 59,* 225-235.

Gagnon, S., Foster, J. K., Turcotte, J. & Jongenelis, S. (2004). Involvement of the hippocampus in implicit learning of supra-span sequences: The case of SJ. *Cognitive Neuropsychology, 21,* 867-882.

Gathercole, S. E. (1995). Is nonword repetition a test of phonological memory or long-term knowledge? It all depends on the nonwords. *Memory and Cognition, 23,* 83-94.

Gathercole, S. E. (1997). Models of verbal short-term memory. In M. A. Conway (Ed.) *Cognitive Models of Memory.* (pp. 13-45). Hove, UK: Psychology Press Publishers.

Gathercole, S. E. (1998). The development of memory. *Journal of Child Psychology and Psychiatry, 39,* 3-27.

Gathercole, S. E. (2006a). Nonword repetition and word learning: The nature of the relationship. *Applied Psycholinguistics, 27,* 513-543.

Gathercole, S. E. (2000b). Complexities and constraints in nonword repetition and word learning. *Applied Psycholinguistics, 27,* 599-613.

Gathercole, S. E. & Adams, A-M. (1993). Phonological working memory in very young children. *Developmental Psychology, 29,* 770-778.

Gathercole, S. E. & Adams, A-M. (1994). Children's phonological working memory: Contributions of long-term knowledge and rehearsal. *Journal of Memory and Language, 33,* 672-688.

Gathercole, S. E., Adams, A-M. & Hitch, G. J. (1994). Do young children rehearse? An individual-differences analysis. *Memory and Cognition, 22,* 201-207.

Gathercole, S. E. & Baddeley, A. D. (1989a). Evaluation of the role of phonological STM in the development of vocabulary in children: A longitudinal study. *Journal of Memory and Language, 28,* 200-213.

Gathercole, S. E. & Baddeley, A. D. (1989b). The role of phonological memory in normal and disordered language development. In C. von Euler, I. Lundberg & G. Lennerstrand (Eds.) *Brain and Reading* (pp. 245-255). London: The Macmillan Press, Ltd.

Gathercole, S. E. & Baddeley, A. D. (1990a). Phonological memory deficits in language disordered children: Is there a causal connection? *Journal of Memory and Language, 29,* 336-360.

Gathercole, S. E. & Baddeley, A. D. (1990b). The role of phonological memory in vocabulary acquisition: A study of young children learning new names. *British Journal of Psychology, 81,* 439-454.

Gathercole, S. E. & Baddeley, A. D. (1993). *Working memory and language*. Hillsdale, NJ: Erlbaum.

Gathercole, S. E. & Baddeley, A. D. (1997). Sense and sensitivity in phonological memory and vocabulary development: A reply to Bowey (1996). *Journal of Experimental Child Psychology, 67,* 290-294.

Gathercole, S. E., Frankish, C. R., Pickering, S. J. & Peaker, S. H. (1999). Phonotactic influences on short-term memory. *Journal of Experimental Psychology: Learning, Memory and Cognition, 25,* 1-13.

Gathercole, S. E. & Hitch, G. J. (1993). Developmental changes in short-term memory: A revised working memory perspective. In A. F. Collins, S. E. Gathercole, M. A. Conway & P. E. Morris (Eds.) *Theories of memory* (pp. 189-209). Hove, UK: Lawrence Erlbaum Associates Ltd.

Gathercole, S. E., Hitch, G. J., Service, E. & Martin, A. J. (1997). Phonological short-term memory and new word learning in children. *Developmental Psychology, 33,* 966-979.

Gathercole, S. E. & Martin, A. J. (1996). Interactive processes in phonological memory. In S. E. Gathercole (Ed.) *Models of short-term memory* (pp. 73-100). Hove, Sussex: Psychology Press.

Gathercole, S. E., Martin, A. J. & Hitch, G. J. (1996). *Wordlikeness effects in the immediate repetition and long-term learning of new words.* Unpublished manuscript.

Gathercole, S. E., Pickering, S. J., Hall, M. & Peaker, S. M. (2001). Dissociable lexical and phonological influences on serial recognition and serial recall. *Quarterly Journal of Experimental Psychology, 54A,* 1-30.

Gathercole, S. E., Service, E., Hitch, G. J., Adams, A-M & Martin, A. J. (1999). Phonological short-term memory and vocabulary development: Further evidence on the nature of the relationship. *Applied Cognitive Psychology, 13,* 65-77.

Gathercole, S. E. & Thorn, A. S. C. (1998). Phonological short-term memory and foreign language learning. In A. F. Healy & L. E. Bourne, Jr (Eds.) *Foreign language learning: Psycholinguistic studies on training and retention* (pp. 141-158). New Jersey: Lawrence Erlbaum Associates, Inc.

Gathercole, S. E., Tiffany, C., Briscoe, J., Thorn, A. & The ALSPAC team. (2005). Developmental consequences of poor phonological short-term memory function in childhood: A longitudinal study. *Journal of Child Psychology and Psychiatry, 46,* 598-611.

Gathercole, S. E., Willis, C. & Baddeley, A. D. (1991b). Nonword repetition, phonological memory and vocabulary: A reply to Snowling, Chiat and Hulme. *Applied Psycholinguistics, 12,* 375-379.

Gathercole, S. E., Willis, C. S., Baddeley, A. D. & Emslie, H. (1994). The Children's Test of Nonword Repetition: A test of phonological working memory. *Memory, 2,* 103-127.

Gathercole, S. E., Willis, C., Emslie, H. & Baddeley, A. D. (1991a). The influences of number of syllables and wordlikeness on children's repetition of nonwords. *Applied Psycholinguistics, 12,* 349-367.

Gathercole, S. E., Willis, C. S., Emslie, H. & Baddeley, A. D. (1992). Phonological memory and vocabulary development during the early school years: A longitudinal study. *Developmental Psychology, 28,* 887-898.

Glasspool, D. W. (1995). Competitive queuing and the articulatory loop: An extended network model. In J. Levy, D. Bairaktaris, J. Bullinaria & D. Cairns (Eds.) *Connectionist models of memory and language.* London: UCL Press.

Gleitman, L. (1993). The structural sources of verb meanings. In P. Bloom (Ed.) *Core readings in language acquisition.* (pp. 174-221). Cambridge, England: Harvester Wheatsheaf.

Grant, J., Kormiloff-Smith, A., Gathercole, S. E., Paterson, S., Howlin, P., Davies, M. & Udwin, O. (1997). Phonological short-term memory and its relationship to language in Williams Syndrome. *Cognitive Neuropsychiatry, 2,* 81-99.

Greg, V. H., Freedman, C. M. & Smith, D. K. (1989). Word frequency, articulatory suppression and memory span. *British Journal of Psychology, 80,* 363-374.

Gupta, P. (2003). Examining the relationship between word learning, nonword repetition and immediate serial recall in adults. *Quarterly Journal of Experimental Psychology, 56A,* 1213-1236.

Gupta, P. (2005). Primacy and recency in nonword repetition. *Memory, 13,* 318-324.

Gupta, P. (2006a). Nonword repetition, phonological storage and multiple determination. *Applied Psycholinguistics, 27,* 545-598.

Gupta, P. (2006b). A computational model of nonword repetition, immediate serial recall and nonword learning. In A. Thorn & M. Page (Eds.). *Interactions between short-term and long-term memory in the verbal domain.* Hove: Psychology Press.

Gupta, P. & MacWhinney, B. (1995). Is the articulatory loop articulatory or auditory? Reexamining the effects of concurrent articulation on immediate serial recall. *Journal of Memory and Language, 34,* 63-88.

Gupta, P. & MacWhinney, B. (1997). Vocabulary acquisition and verbal short-term memory: computational and neural bases. *Brain and Language, 59,* 267-333.

Halliday, M. S., Hitch, G. J., Lennon, B. & Pettifer, C. (1990). Verbal short-term memory in children: The role of the articulatory loop. *European Journal of Cognitive Psychology, 2,* 23-38.

Hartley, T. & Houghton, G. (1996). A linguistically constrained model of short-term memory for nonwords. *Journal of Memory and Language, 35,* 1-31.

Hebb, D. O. (1949). *Organisation of behaviour.* New York: Wiley.

Hebb, D. O. (1961). Distinctive features of learning in the higher animal. In J. F. Delafresnaye (Ed.) *Brain mechanisms and learning* (pp. 37-51). Oxford: Blackwell Scientific Press.

Henry, L. A. (1991). The effect of word length and phonemic similarity in young children's short-term memory. *Quarterly Journal of Experimental Psychology, 43A,* 35-52.

Henson, R. N. A. (1998). Short-term memory for serial order: the start-end model. *Cognitive Psychology, 36,* 73-137.

Henson, R. N. A. (1999). Positional information in short-term memory: Relative or absolute? *Memory and Cognition, 27,* 915-927.

Henson, R. N. A. (2001). Serial order for short-term memory. *The Psychologist, 14,* 70-73.

Henson, R. N. A., Hartley, T., Burgess, N., Hitch, G. J. & Flude, B. (2003). Selective interference with verbal 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. A., 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.

Heron, A. & Craik, F. (1964). Age differences in cumulative learning of meaningful and meaningless material. *Scandinavian Journal of Psychology, 5,* 209-217.

Hitch, G. J., Fastame, M. C. & Flude, B. (2005). How is the serial order of a verbal sequence coded? Some comparisons between models. *Memory, 13,* 247-258.

Hitch, G. J., Flude, B. & Burgess, N. (in press). Slave to the rhythm: A mechanism for short-term memory and long-term learning of verbal sequences. *Journal of Memory and Language.*

Hitch, G. J. & Halliday, M. S. (1983). Working memory in children. *Philosophical Transactions of the Royal Society, Series B, 302*, 324-340.

Hitch, G. J., Halliday, M. S., Dodd, A. & Littler, J. E. (1989a). Development of rehearsal in short-term memory: Differences between pictorial and spoken stimuli. *British Journal of Developmental Psychology, 7,* 347-362.

Hitch, G. J., Halliday, M. S. & Littler, J. E. (1989b). Item identification time and rehearsal rate as predictors of memory span in children. *Quarterly Journal of Experimental Psychology, 41A,* 321-328.

Hitch, G. J., Halliday, M. S., Schaafstal, A. M. & Heffernan, T. M. (1991). Speech, "inner speech", and the development of short-term memory: Effects of picture-labelling on recall. *Journal of Experimental Child Psychology, 51,* 220-234.

Hitch, G. J., Halliday, M. S., Schaafstal, A. M., & Schraagen, J. M. C. (1988). Visual working memory in young children. *Memory and Cognition, 16,* 120-132.

Hitch, G. J., McNeil, A. M., Page, M. P. A., Cumming, N. & Norris, D. G. (2006). *Do children show a Hebb effect?* Unpublished manuscript.

Houghton, G. (1990). The problem of serial order: A neural network memory of sequence learning and recall. In R. Dale, C. Mellish & M. Zock (Eds). *Current research in natural language generation.* London: Academic Press.

Howell, D. C. (1997). *Statistical methods for Psychology.* (4th Edition). London, England: Wadsworth Publishing Company.

Hulme, C. (1987). The effects of acoustic similarity on memory in children: A comparison between visual and auditory presentation. *Applied Cognitive Psychology, 1,* 45-52.

Hulme, C., Maughan, S. & Brown, G. D. A. (1991). Memory for familiar and unfamiliar words: Evidence for a long-term memory contribution to short-term memory span. *Journal of Memory and Language, 30,* 685-701.

Hulme, C., Roodenrys, S., Brown, G. D. A. & Mercer, R. (1995). The role of long-term memory mechanisms in memory span. *British Journal of Psychology, 86,* 527-536.

Hulme, C., Roodenrys, S., Schweickert, R., Brown, G. D. A., Martin, S. & Stuart, G. (1997). Word-frequency effects on short-term memory tasks: Evidence for a redintegration process in immediate serial recall. *Journal of Experimental Psychology: Learning, Memory and Cognition, 23,* 1217-1232.

Hulme, C., Thomson, N., Muir, C. & Lawrence, A. (1984). Speech rate and the development of short-term memory span. *Journal of Experimental Child Psychology, 38,* 241-253.

Hulme, C. & Tordoff, V. (1989). Working memory development: The effects of speech rate, word length and acoustic similarity on serial recall. *Journal of Experimental Child Psychology, 47,* 72-87.

Jones, D. M. (1993). Objects, streams and threads of auditory attention. In A. D. Baddeley & L. Weiskrantz (Eds.). *Attention: Selection, awareness and control* (pp. 87-103). Oxford, England: Clarendon Press.

Jones, D. M., Beaman, P. & Macken, W. J. (1996). The object-orientated episodic record model. In S. E. Gathercole (Ed.) *Models of short-term memory.* (pp. 209-238). Hove, Sussex: Psychology Press.

Jones, D. M., Farrand, P., Stuart, G. & Morris, N. (1995). Functional equivalence of verbal and spatial information in serial short-term memory. *Journal of Experimental Psychology: Learning, Memory and Cognition, 21,* 1008-1018.

Jones, D. M. & Hughes, R. W. & Macken, W. J. (2006). Perceptual organisation masquerading as phonological storage: Further support for a perceptual-gestural view of short-term memory. *Journal of Memory and Language, 54,* 265-281.

Jones, D. M. & Hughes, R. W. & Macken, W. J. (2007). Commentary on Baddeley and Larsen (2007). The phonological store abandoned. *Quarterly Journal of Experimental Psychology, 60,* 505-511.

Jones, D. M. & Macken, W. J. (1993). Irrelevant tones produce an irrelevant speech effect: Implications for phonological coding in working memory. *Journal of Experimental Psychology: Learning, Memory and Cognition, 19,* 369-381.

Jones, D. M. & Macken, W. J. (1995). Phonological similarity in the irrelevant speech effect. Within- or between-stream similarity? *Journal of Experimental Psychology: Learning, Memory and Cognition, 21,* 103-115.

Jones, D. M., Macken, W. J. & Nicholls, A. P. (2004). The phonological store of working memory: Is it phonological and is it a store? *Journal of Experimental Psychology: Learning, Memory and Cognition, 30,* 656-674.

Jusczyk, P. W., Luce, P. A. & Charles-Luce, J. (1994). Infants' sensitivity to phonotactic patterns in the native language. *Journal of Memory and Language, 33,* 630-645.

Karlsen, P. J., Imenes, A. G., Johannessen, K., Endestad, T. & Lian, A. (2007). Why does the phonological similarity effect reverse with nonwords? *Psychological Research, 71,* 448-457.

Karlsen, P. J. & Lian, A. (2005). Modulating the phonological similarity effect: The contribution of interlist similarity and lexicality. *Memory and Cognition, 33,* 542-556.

Kidd, G. R. & Greenwald, A. G. (1988). Attention, rehearsal and memory for serial order. *American Journal of Psychology, 101,* 259-279.

Kintsch, W. (1977). *Memory and cognition.* New York: John Wiley & Sons, Inc.

Kucera, H. & Francis, W. M. (1967). *Computational analysis of present-day American English.* Providence, RI: Brown University Press.

Larsen, J. D. & Baddeley, A. D. (2003). Disruption of verbal STM by irrelevant speech, articulatory suppression and manual tapping: Do they have a common source? *Quarterly Journal of Experimental Psychology, 56A,* 1249-1268.

Larsen, J. D., Baddeley, A. D. & Andrade, J. (2000). Phonological similarity and the irrelevant speech effect: Implications for models of short-term verbal memory. *Memory, 8,* 145-157.

Lee, C. L. & Estes, W. K. (1977). Order and position in primary memory for letter strings. *Journal of Verbal Learning and Verbal Behaviour, 16,* 395-418.

Lee, C. L. & Estes, W. K. (1981). Item and order information in short-term memory: Evidence for multilevel perturbation processes. *Journal of Experimental Psychology, 7,* 149-169.

Leonard, L. (1998). *Children with specific language impairments.* Cambridge, MA: MIT Press.

Levy, B. A. (1971). The role of articulation in auditory and visual short-term memory. *Journal of Verbal Learning and Verbal Behaviour, 10,* 123-132.

Lewandowsky, S. & Murdock, B. B. (1989). Memory for serial order. *Psychological Review, 96,* 25-57.

Lian, A. & Karlsen, P. J. (2004). Advantages and disadvantages of phonological similarity in serial recall and serial recognition of nonwords. *Memory and Cognition, 32,* 223-234.

Lian, A., Karlsen, P. J. & Eriksen, T. B. (2004). Opposing effects of phonological similarity on item and order memory of words and nonwords in the serial recall task. *Memory, 12,* 314-337.

Lian, A., Karlsen, P. J. & Winsvold, B. (2001). A re-evaluation of the phonological similarity effect in adults' short-term memory of words and nonwords. *Memory, 9,* 281-299.

Logie, R. H. (1995). *Visuo-spatial working memory.* Hove, Sussex: Lawrence Erlbaum Associates, Ltd.

Logie, R. H., Della Sala, S., Laiacona, M., Chalmers, P. & Wynn, V. (1996). Group aggregates and individual reliability: The case of verbal short-term memory. *Memory and Cognition, 24,* 305-321.

Longoni, A. M., Richardson, A. T. E. & Aiello, A. (1993). Articulatory rehearsal and phonological storage in working memory. *Memory and Cognition, 21,* 11-22.

Lovatt, P., Avons, S. E. & Masterson, J. (2000). The word length effect and disyllabic words. *Quarterly Journal of Experimental Psychology, 53A,* 1-22.

Lovatt, P., Avons, S. E. & Masterson, J. (2002). Output decay immediate serial recall: Speech time revisited. *Journal of Memory and Language, 46,* 227-243.

Markman, E. M. (1994). Constraints on word meaning in early language acquisition. *Lingua, 92,* 199-227.

Markman, E. M. & Wachtel, G. F. (1988). Children's use of mutual exclusivity to constrain the meanings of words. *Cognitive Psychology, 20,* 121-157.

Markman, E. M., Wasow, J. L. & Hansen, M. B. (2003). Use of mutual exclusivity assumption by young word learners. *Cognitive Psychology, 47,* 241-275.

Martin, A. & Gathercole, S. E. (1997). *Phonological neighbourhood analysis of nonwords as a function of wordlikeness.* Unpublished manuscript.

Masoura, E. V. & Gathercole, S. E. (1999). Phonological short-term memory and foreign language learning. *International Journal of Psychology, 34,* 383-388.

Masoura, E. V. & Gathercole, S. E. (2005). Contrasting contributions of phonological short-term memory and long-term knowledge to vocabulary learning in a foreign language. *Memory, 13,* 422-429.

McKelvie, S. J. (1987). Learning and awareness in the Hebb digits task. *Journal of General Psychology, 114,* 75-88.

Melton, A. W. (1963). Implications of short-term memory for a general theory of memory. *Journal of Verbal Learning and Verbal Behaviour, 2,* 1-21.

Metsala, J. L. (1999). Young children's phonological awareness and nonword repetition as a function of vocabulary development. *Journal of Experimental Psychology, 91,* 3-19.

Metsala, J. K. & Walley, A. C. (1998). Spoken vocabulary growth and the segmental restructuring of lexical representations: Precursors to phonemic awareness and early reading ability. In J. L. Metsala & L. C. Ehri (Eds.) *Word recognition in beginning literacy* (pp. 89-120). Mahwah, NJ: Erlbaum.

Michas, I. C. & Henry, L. A. (1994). The link between phonological memory and vocabulary acquisition. *British Journal of Developmental Psychology, 12,* 147-163.

Milberg, W., Alexander, M. P., Charness, N., McGlinchey-Berroth, R. & Barrett. A. (1988). Learning of a complex arithmetic skill in amnesia: Evidence for a dissociation between compilation and production. *Brain and Cognition, 8,* 91-104

Montgomery, J. (1995). Examination of phonological working memory in specifically language-impaired children. *Applied Psycholinguistics, 16,* 355-378.

Mueller, S. T., Seymour, T. L., Kieras, D. E. & Meyer, D. E. (2003). Theoretical implications of articulatory duration, phonological similarity, and phonological complexity in verbal working memory. *Journal of Experimental Psychology: Learning, Memory and Cognition, 29,* 1353-1380.

Munson, B., Edwards, J. & Beckman, M. E. (2005). Relationships between nonword repetition accuracy and other measures of linguistic development in children with phonological disorders. *Journal of Speech, Language and Hearing Research, 48,* 61-78.

Murray, D. J. (1967). The role of speech responses in short-term memory. *Canadian Journal of Psychology, 21,* 263-276.

Murray, D. J. (1968). Articulation and acoustic confusability in short-term memory. *Journal of Experimental Psychology, 78,* 679-684.

Nairne, J. S. & Kelley, M. R. (1999). Reversing the phonological similarity effect. *Memory and Cognition, 27,* 45-53.

Neath, I., Brown, G. D. A., Poirier, M. & Fortin, C. (2005). Short-term and working memory: Past, progress and prospects. *Memory, 13,* 225-235.

Nicolson, R. (1981). The relationship between memory span and processing speed. In M. Friedman, J. P. Das & N. O'Conner (Eds). *Intelligence and learning* (pp. 179-184). New York: Plenum Press.

Nijland, L., Maassen, B., van der Meulen, S., Gabreels, F., Kraaimaat, F. W. & Schreuder, R. (2002). Coarticulation patterns in children with developmental apraxia of speech. *Clinical Linguistics and Phonetics, 16,* 461-483.

Nimmo, L. M. & Roodenrys, S. (2005). The phonological similarity effect in serial recognition. *Memory, 13,* 773-784.

Page, M. P. A., Cumming, N., Norris, D. G., Hitch, G. J. & McNeil, A. M. (2006). Repetition learning in the immediate serial recall of visual and auditory materials. *Journal of Experimental Psychology: Learning, Memory and Cognition, 32,* 716-733.

Page, M. P. A. & Norris, D. G. (1998). The primacy model: a new model of immediate serial order recall. *Psychological Review, 105,* 761-781.

Papagno, C., Valentine, T. & Baddeley, A. D. (1991). Phonological short-term memory and foreign-language vocabulary learning. *Journal of Memory and Language, 30,* 331-347.

Papagno, C. & Vallar, G. (1992). Phonological short-term memory and the learning of novel words: The effect of phonological similarity and item length. *Quarterly Journal of Experimental Psychology, 44A,* 47-67.

Papagno, C. & Vallar, G. (1995). Verbal short-term memory and vocabulary learning in polyglots. *Quarterly Journal of Experimental Psychology, 48A,* 98-107.

Peterson, L. R. (1966). Short-term verbal memory and learning. *Psychological Review, 73,* 193-207.

Peterson, L. R. & Johnson, S. T. (1971). Some effects of minimising articulation on short-term retention. *Journal of Verbal Learning and Verbal Behaviour, 10,* 346-354.

Pinker, S. (1984). *Language learnability and language development.* Cambridge, MA: Harvard University Press.

Posner, M. I. & Konick, A. F. (1966). On the role of interference in short-term retention. *Journal of Experimental Psychology, 72,* 221-231.

Rooendrys, S. & Hinton, M. (2002). Sublexical or lexical effects on serial recall of nonwords? *Journal of Experimental Psychology: Learning, Memory and Cognition, 28,* 29-33.

Roodenrys, S., Hulme, C., Alban, J., Ellis, A. & Brown, G. D. A. (1994). Effects of word frequency and age of acquisition on short-term memory span. *Memory and Cognition, 22,* 695-701.

Roodenrys, S., Hulme, C., Lethbridge, A., Hinton, M. & Nimmo, L. M. (2002). Word-frequency and phonological-neighbourhood effects on verbal short-term memory. *Journal of Experimental Psychology: Learning, memory and Cognition, 28,* 1019-1034.

Roy, P. & Chiat, S. (2004). A prosodically controlled word and nonword repetition task for 2- to 4-year-olds: Evidence from typically developing children. *Journal of Speech, Language and Hearing Research, 47,* 223-234.

Saffran, J. R., Newport, E. L. & Aslin, R. N. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language, 35,* 606-621.

Saint-Aubin, J. & Poirier, M. (1999). Semantic similarity and immediate serial recall: Is there a detrimental effect on order information? *Quarterly Journal of Experimental Psychology, 52A,* 367-394.

Salamé, P. & Baddeley, A. D. (1982). Disruption of memory by unattended speech: Implications for the structure of working memory. *Journal of Verbal Learning and Verbal Behaviour, 21,* 150-164.

Salamé, P. & Baddeley, A. D. (1986). Phonological factors in STM: Similarity and the unattended speech effect. *Bulletin of the Psychonomic Society, 24,* 263-265.

Schwartz, M. & Bryden, M. P. (1971). Coding factors in the learning of repeated digit sequences. *Journal of Experimental Psychology, 87,* 331-334.

Sechler, E. S. & Watkins, M. J. (1991). Learning to reproduce a list and memory for the learning. *American Journal of Psychology, 104,* 367-394.

Seger, C. A. (1994). Implicit learning. *Psychological Bulletin, 115,* 163-196.

Service, E. (1992). Phonology, working memory and foreign-language learning. *Quarterly Journal of Experimental Psychology, 45A,* 21-50.

Service, E. (1998). The effect of word length on immediate serial recall depends on phonological complexity, not articulatory duration. *Quarterly Journal of Experimental Psychology, 51A,* 283-304.

Service, E. (2000). Phonological complexity and word duration in immediate recall: Different paradigms answer different questions. A comment on Cowan, Nugent, Elliot and Geer. *Quarterly Journal of Experimental Psychology, 53A,* 661-665.

Service, E. & Craik, F. (1993). Differences between young and older adults in learning a foreign vocabulary. *Journal of Memory and Language, 32,* 608-623.

Service, E. & Kohonen, V. (1995). Is the relation between phonological memory and foreign language learning accounted for by vocabulary acquisition? *Applied Psycholinguistics, 16,* 155-172.

Shallice, T. & Warrington, E. K. (1970). Independent functioning of verbal memory stores: A neuropsychological study. *Quarterly Journal of Experimental Psychology, 22,* 261-273.

Smith, M. E. (1926). An investigation of the development of the sentence and the extent of vocabulary in young children. *University of Iowa Studies in Child Welfare, 3(5).*

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., Hay, D. C., Hitch, G. J. & Horton, N. J. (2005). Serial position memory in the visual-spatial domain: Reconstructing sequences of familiar faces. *Quarterly Journal of Experimental Psychology, 58A,* 909-930.

Snowling, M., Chiat, S. & Hulme, C. (1991). Words, nonwords and phonological processes: Some comments on Gathercole, Willis, Emslie and Baddeley. *Applied Psycholinguistics, 12,* 369-373.

Snowling, M., Goulandris, A., Bowlby, M. & Howell, P. (1986). Segmentation and speech perception in relation to reading skill: A developmental analysis. *Journal of Experimental Child Psychology, 41,* 489-507.

Snowling, M. & Hulme, C. (1989). A longitudinal case study of developmental phonological dyslexia. *Cognitive Neuropsychology, 6,* 379-401.

Stadler, M. A. (1993). Implicit serial learning: Questions inspired by Hebb (1961). *Memory and Cognition, 21,* 819-827.

Stark, R. E. & Tallal, P. (1981). Selection of children with specific language deficits. *Journal of Speech and Hearing Disorders, 49,* 114-122.

Sternberg, R. (1987). Most vocabulary is learned from context. In M. McKeown & M. Curtis (Eds.) *The nature of vocabulary acquisition* (pp. 89-106). Hillsdale, NJ: Erlbaum.

Tallal, P. & Piercy, M. (1975). Developmental aphasia: The perception of brief vowels and extended stop consonants. *Neuropsychologia, 12,* 83-94.

Tallal, P., Stark, R. E. & Mellitts, D. (1985). The relationship between auditory temporal analysis and receptive language development: Evidence from studies of developmental language disorder. *Neuropsychologia, 23,* 527-534.

Tehan, G. & Humphreys, M. S. (1988). Articulatory loop explanations of memory span and pronunciation rate correspondences: A cautionary note. *Bulletin of the Psychonomic Society, 26,* 293-296.

Thorn, A. S. C. & Frankish, C. R. (2005). Long-term knowledge effects on serial recall of nonwords are not exclusively lexical. *Journal of Experimental Psychology: Learning, Memory and Cognition, 31,* 729-735.

Thorn, A. S. C. & Gathercole, S. E. (1999). Language-specific knowledge and short-term memory in bilingual and non-bilingual children. *Quarterly Journal of Experimental Psychology, 52A,* 303-324.

Thorn, A. S. C. & Gathercole, S. E. (2001). Language differences in verbal short-term memory do not exclusively originate in the process of subvocal rehearsal. *Psychonomic Bulletin and Review, 8,* 357-364.

Thorn, A. S. C., Gathercole, S. E. & Frankish, C. R. (2002). Language familiarity effects in short-term memory: The role of output delay and long-term knowledge. *Quarterly Journal of Experimental Psychology, 55A,* 1363-1383.

Thorn, A. S. C., Gathercole, S. E. & Frankish, C. R. (2005). Redintegration and the benefits of long-term knowledge in verbal short-term memory: An evaluation of Schweickert's (1993) multinomial processing tree model. *Cognitive Psychology, 50,* 133-158.

Trojano, L. & Grossi, D. (1995). Phonological and lexical coding in verbal short-term memory and learning. *Brain and Language, 51,* 336-354.

Turcotte, J., Gagnon, S. & Poirier, M. (2005). The effect of old age on the learning of supraspan sequences. *Psychology and Aging, 20,* 251-260.

Vallar, G. & Baddeley, A. D. (1984a). Fractionation of working memory: Neuropsychological evidence for a phonological short-term store. *Journal of Verbal Learning and Verbal Behaviour, 23,* 151-161.

Vallar, G. & Baddeley, A. D. (1984b). Phonological short-term store, phonological processing and sentence comprehension: A neuropsychological case study. *Cognitive Neuropsychology, 1,* 121-141.

Vallar, G. & Papagno, C. (2002). Neuropsychological impairments of verbal short-term memory. In A. D. Baddeley, M. D. Kopelman & B. A Wilson (Eds) *The Handbook of Memory Disorders* (2nd Edition) (pp. 249-270). West Sussex, England: John Wiley & Sons Ltd.

Vallar, G. & Shallice, T. (Eds.) (1990). *Neuropsychological impairments of short-term memory.* Cambridge, England: Cambridge University Press.

van Bon, W. H. J. & van der Pijl, J. M. L. (1997). Effects of word length and wordlikeness on pseudoword repetition by poor and normal readers. *Applied Psycholingustics, 18,* 101-114.

van der Lely, H. K. J. & Howard, D. (1993). Children with specific language impairment: Linguistic impairment of short-term memory deficit? *Journal of Speech and Hearing Research, 36,* 1193-1207.

Vitevitch, M. S., Luce, P. A., Charles-Luce, J. & Kemmerer, D. (1997). Phonotactics and syllable stress: Implications for the processing of spoken nonsense words. *Language and Speech, 40,* 47-62.

Vitevitch, M. S., Luce, P. A., Pisoni, D. B. & Auer, E. T. (1999). Phonotactics, neighbourhood activation and lexical access for spoken words. *Brain and Language, 68,* 306-311.

Walley, A. C. (1993). The role of vocabulary development in children's spoken word recognition and segmentation ability. *Developmental Review, 13,* 286-350.

Ward, G., Avons, S. E. & Melling, L. (2005). Serial position curves in short-term memory: Functional equivalence across modalities. *Memory, 13,* 308-317.

Watkins, M. J. (1977). The intricacy of memory span. *Memory and Cognition, 5,* 529-534.

Wickelgren, W. A. (1965a). Short-term memory for phonemically similar lists. *American Journal of Psychology, 78,* 567-574.

Wickelgren, W. A. (1965b). Acoustic similarity and intrusion errors in short-term memory. *Journal of Experimental Psychology, 70,* 102-108.

Wilson, B. A. (2002). Assessment of memory disorders. In A. D. Baddeley, M. D. Kopelman & B. A Wilson (Eds) *The Handbook of Memory Disorders* (2nd Edition) (pp. 617-636). West Sussex, England: John Wiley & Sons Ltd.

Wright, C. C. (1979). Duration differences between rare and common words and their implications for the interpretation of word frequency effects. *Memory and Cognition, 7,* 411-419.