Published in final edited form as:

Aphasiology. 2012 January 1; 26(3-4): . doi:10.1080/02687038.2011.642795.

Slave systems in verbal short-term memory

David Caplan¹, Gloria Waters², and David Howard³

¹Department of Neurology, Massachusetts General Hospital, Boston, MA, USA

²Department of Speech and Hearing Sciences, Sargent College, Boston University, Boston, MA, USA

³School of Education, Communication and Language Sciences, University of Newcastle upon Tyne, Newcastle upon Tyne, UK

Abstract

Background—The model of performance in short-term memory (STM) tasks that has been most influential in cognitive neuropsychological work on deficits of STM is the “working memory” model mainly associated with the work of Alan Baddeley and his colleagues.

Aim—This paper reviews the model. We examine the development of this theory in studies that account for STM performances in normal (non-brain-damaged) individuals, and then review the application of this theory to neuropsychological cases and specifications, modifications, and extensions of the theory that have been suggested on the basis of these cases. Our approach is to identify the major phenomena that have been discussed and to examine selected papers dealing with those phenomena in some detail.

Main Contribution—The main contribution is a review of the WM model that includes both normative and neuropsychological data.

Conclusions—We conclude that the WM model has many inconsistencies and empirical inadequacies, and that cognitive neuropsychologists might benefit from considering other models when they attempt to describe and explain patients’ performances on STM tasks.

Keywords

Aphasia; Aphasiology; Neuropsychology; Psychosocial; Language; Brain

This paper discusses the two “slave systems” in the working memory model of verbal short memory developed by Alan Baddeley and his colleagues: the phonological store (PS) and an articulatory rehearsal process. It reviews the evidence for these aspects of the model from studies of neurologically normal individuals and neuropsychological cases. We argue that evidence for these slave systems is weak.

We start with some terminology. Baddeley’s model is widely known as the “working memory (WM)” model, because it was first proposed by Baddeley and Hitch (1974) as a model of the combined ability to maintain small amounts of information in an active, available state for a short period of time (i.e., of STM) and to perform computations on that information in the service of a task. The theoretical heart of that model was the “central executive” (CE), a component of memory that maintained abstract representations and had a

computational function; visuo-spatial and verbal “slave systems” were recruited to aid with the maintenance function when the CE was overloaded. However, Baddeley (2000b) separated the computational and memory functions of the CE, and added to his model an “episodic buffer” (EB)—an amodal, or perhaps multimodal, store—that has purely memory functions. The term “working memory” (WM) model is now widely used to refer to the slave systems.

The verbal slave system consists of two components: a “phonological store (PS)” that maintains information in phonological form subject to rapid decay, and a rehearsal mechanism that can be used to retain information in the PS over longer time intervals. These two components were together originally called the “articulatory loop” (AL), a term that was replaced by the term “phonological loop” in later work. Evidence for a role of phonological representations in STM came from the finding that phonological but not semantic or orthographic similarity led to confusability of items in auditory immediate serial recall—the “phonological similarity effect” (PSE) (Baddeley, 1966a, 1966b; Conrad, 1963, 1964; Wickelgren, 1965). A variety of controls in studies by Baddeley and others (Colle & Welsh, 1976; Salame & Baddeley, 1982; Sperling 1967) argued strongly that the PS was not an echoic (iconic), or pre-categorical, store. Evidence for rapid decay of items in STM came in part from effects of filled and unfilled delays in immediate serial recall and free recall (Brown, 1958; Glanzer & Cunitz, 1966; Peterson & Peterson, 1959) and in part from the word length effect (WLE), the finding that span is larger for short than for long words (Baddeley, Thomson, & Buchanan, 1975). Baddeley et al. (1975) claimed that the WLE was a function of the articulatory duration of words, not their phonological complexity, consistent with the view that representations decayed in a STM system, where the amount of *time* needed for rehearsal was what determined performance. Baddeley et al. (1975) also found that the WLE was correlated with articulation rate; participants could repeat the number of words that they could articulate in about 2 seconds. This observation led to the suggestion that an articulatory-based rehearsal process played a role in STM. Evidence that the PSE and WLE were due to separate mechanisms came from the finding that the PSE, but not the WLE, in auditory immediate serial recall persisted under concurrent articulation (Baddeley et al., 1975), pointing to a mechanism other than articulatory-based rehearsal supporting maintenance of phonological representations, and that concurrent irrelevant speech reduced visual span to the same extent regardless of the length of the irrelevant words, suggesting that irrelevant speech did not reduce visual span by “pre-empting” an articulatory process (Salame & Baddeley, 1982, p. 157). The findings that the visual PSE and the effect of concurrent unattended speech on visual span were reduced or eliminated by concurrent articulation (Baddeley et al., 1975; Salame & Baddeley, 1982) were consistent with the view that articulatory-based rehearsal helps maintain representations in the PS.

This model is widely (though not universally) endorsed; e.g., Page and Henson (2001) say it “gives a very good, indeed unsurpassed, qualitative account of the data in which we are interested (p. 194)” and predict “a bright future” for the model (p. 195). It has been related to neuropsychological cases and to neuroimaging data, and the slave systems have been said to be important to two language functions that are arguably more ecologically valid than immediate serial recall and other STM tasks: aspects of language comprehension (Vallar & Baddeley, 1987) and vocabulary learning (Baddeley, Gathercole, & Papagno, 1998).

THE WM MODEL: NORMATIVE DATA

We will review four empirical phenomena that play critical roles in Baddeley’s model—the word length effect; the effect of concurrent articulation; the phonological similarity effect; and the irrelevant speech effect—and one that is less central to the model but that is important in STM; serial position effects, especially recency. The first two effects are

primarily related to the role of an articulatory mechanism in STM. The second two are mainly related to the role of phonological representations (the PS) in STM. In both pairs the first phenomenon represents the positive phenomenon that the STM component is said to account for, and the second represents the effect of interfering with that component. The last set of phenomena serves to expand the discussion to phenomena other than those that the WM model considers.

The word length effect

Duration vs capacity effect of length—Baddeley et al. (1975) studied the WLE effect in an effort to consider the issue of “chunking” in memory. Previous studies showing an effect of word length on span seemed to contradict the view that single words were chunks and that span was capacity (chunk) limited (Miller, 1956). However, this result could be explained if chunks were sublexical units such as syllables. Baddeley et al. therefore undertook a series of studies to determine whether two sublexical features—number of phonemes and syllables—and/or the temporal duration of words, determined span. They argued that the length effect was due to temporal, not phonological, factors. The temporal basis of the WLE led to two theoretical conclusions: (1) items in the PS are subject to decay over time, and (2) an articulatory-based process supports span through rehearsal.

A duration-based WLE is important because it has been considered to be one of the major sources of evidence for the view that representations in the PS are subject to decay. For instance, working within a quite different framework, Cowan (1995; see also Ricker, AuBuchon, & Cowan, 2010) argued that evidence for decay of representations in STM is limited and insecure, and often has other explanations; e.g., the Brown-Petersen effect can be due to interference. Cowan cites the WLE as one of the main phenomena supporting decay in STM. The implication of the WLE that items in the PS (or in STM) decay over time critically depends on the WLE being time, not complexity, based. If longer items are also more complex, reduced span could be due to capacity, not temporal, limits on a store. This issue has led to a series of studies of long- and short-duration items in immediate serial recall and recall-from-end, as well as measurement of temporal duration of pauses during responses in recall tasks. In these studies only the original Baddeley et al. (1975) stimuli have been found to produce a time-based word length effect; seven other sets of stimuli that differ in pronunciation time but are otherwise equated do not show the result. We briefly review the data here.

Baddeley et al. (1975) used bisyllabic words that differed in spoken duration which were matched for frequency and number of phonemes (Exp. 3), and for frequency, number of phonemes, and number of syllables (Exp. 4). Caplan, Rochon, and Waters (1992), Service (1998), Toland and Tehan (2005), Zhang and Feng (1990) and Lovatt, Avons, and Masterson (2000) did not replicate this result in English, Finnish, and Chinese. Neath, Bireta, and Surprenant (2003) found that, of four sets of words that differed in length and were equated for number of syllables and phonemes, only those from Baddeley et al. (1975) showed the WLE (for discussion see Jalbert, Neath, Bireta, & Surprenant, 2011; Jalbert, Neath, & Surprenant, 2011). Issues that have arisen in the discussions in these papers include how to measure word duration (in isolation, for selected pairs of words, for all lists used in the study); whether it is important to test the same participants for word duration and recall; and the source of reversed word length effects seen in some studies (Caplan et al., 1992; Caplan & Waters, 1994; Lovatt et al., 2000). A leading candidate for explaining this last finding is a confound between duration and phonological similarity, which leads to the next topic.

Factors confounded and correlated with the WLE—Mueller, Seymour, Kieras, and Meyer (2003) reported a study in which they used new measures of phonological similarity and articulation duration to predict span for nine word lists. Controlling for phonological similarity, an effect of duration, but not complexity, remained significant in regressions of these factors against span performance. Mueller et al. (2003) concluded that there is a duration-related WLE. However, Lewandowsky and Oberauer (2008) pointed out that the model presented by Mueller et al., containing phonological similarity and duration, accounted for an insignificant 1% more variance ($r^2 = .99$) than one containing the log of the number of phonemes and their neighbourhood distribution (the number of letter positions at which neighbours can be formed) ($r^2 = .98$), and both models produced predicted values that fell within 95% confidence intervals of the data.

This result is one of several that point to the possible confound of lexical variables with word length in producing effects in immediate serial recall. Jalbert, Neath, Bireta, et al. (2011) found that word length was confounded with orthographic neighbourhood size in 11 studies showing a WLE. In their studies the WLE for mono- and tri-syllabic words disappeared when long and short words were matched for orthographic neighbourhood size. Thus, even the WLE found with words that differ in both complexity and duration may not show that phonological complexity affects STM function; Jalbert, Neath, Bireta, et al. (2011) conclude that “the word length effect may be better explained by the differences in linguistic and lexical properties of short and long words rather than by length *per se* (p. 338).”

Forwards and backwards recall—Cowan et al. (1992) varied spoken duration of words early and late in mixed lists of five items and required recall in forward and backward order. With forward recall, recall was greater for lists in which short words appeared in the first half. Excluding the last position, hierarchical regression confirmed that inclusion of the spoken duration of each word recalled prior to the currently recalled word improved prediction of percent recalled better than a regression that included the number of intervening words but not their duration (regression with duration: $R^2 = .94$; regression without duration: $R^2 = .81$; partial $R^2 = .12$, $p < .001$). With backwards recall the effect of position of long and short words in the list was reversed, consistent with output response time affecting recall and inconsistent with a rehearsal mechanism. Cowan et al. (1992) take this result as strong evidence for temporal decay of representations in STM during output. However, the effect of delay on the last recalled words was eliminated when trials on which errors on the initial words were removed, suggesting recall failures, not length, led to the effect (for discussion see Lovatt, Avons, & Masterson, 2002). In addition, regression of time against recall fitted the backwards recall data much less well ($R^2 = .83$) than it did the forward recall data in Exp. 3 ($R^2 = .96$), and did not differ from a regression based on number of items ($R^2 = .77$; partial $R^2 = .05$, *ns*). Also, Cowan et al. (1992) used many words drawn from those used by Baddeley et al. (1975), and orthographic neighbourhood size differed for their short words (2.4; $SD = 1.3$) and long words (0.4; $SD = .5$); ($t = 3.1$; $p = .02$). It thus appears that correlated features may account for Cowan’s results.

Mixed lists—Hulme, Surprenant, Bireta, Stuart, and Neath (2004) reported that lists containing only one-syllable words were better recalled than lists containing only five-syllable words, but that long and short words were equally well recalled in mixed lists. They concluded that neither item- nor list-related temporal decay theories could account for this effect, which they attributed to complexity and stimulus distinctiveness. Cowan, Baddeley, Elliott, and Norris (2003) reported different results, finding that recall of mixed lists was poorer than that of pure short word lists and that recall of short words was better than long words in mixed lists. Bireta, Neath, and Surprenant (2006) used each group’s stimuli with

their original list types and with the other group's list types. They replicated both sets of results with the original materials, regardless of list composition. They also found the Hulme et al. pattern with a new set of words. They concluded that the Cowan et al. result is item specific. A subset of the Cowan et al. items were from Baddeley et al. (1975); Biereta et al. suggested the relevant factor might be imageability of Cowan et al.'s long and short items.

WLE during recall—It remains possible that a duration-based WLE occurs during response production. In a study of 4- and 8-year-olds, Cowan et al. (1992) found that word length affected the duration of words in the response but it had no effect on the duration of inter-response intervals; however, age affected the duration of inter-response intervals (for findings in adults, see Hulme, Newton, Cowan, Stuart, & Brown, 1999). On the basis of such findings, Cowan and Kail (1996) suggested the processes that occur during inter-response intervals might not include covert rehearsal, but rather consist of search. Using a different approach, Cowan, Nugent, Elliott, and Geer (2000) required participants to produce words slowly or quickly, and found an effect of spoken duration on span, a result that they argued established a role for duration-based effects on span during response production. However, subsequent studies (Cowan et al., 2006) found the duration effect disappeared when the response deadline that had been used by Cowan et al. (2000) was eliminated. Thus it is likely that the duration effect was a result of the use of a deadline-response methodology, in which faster responses left more time for search/retrieval to occur (Lewandowky & Oberauer, 2008). There is at present no strong evidence that articulatory-based rehearsal occurs during inter-response intervals. The sole locus of a role for articulation derived from the WLE appears to be on spoken output duration, a virtual necessity of no interest to theories of STM.

Conclusions concerning the WLE—The evidence for a duration-based WLE is based on one set of words. There is no evidence that duration-based WLE affects pauses during recall of items during response production. Non-duration-based WLE effects appear to be due to lexical properties of words, not their phonological complexity. We conclude that the WLE effect provides no support for either a temporally decaying store of phonological representations or an articulatory rehearsal process in STM.

The effect of concurrent articulation

Concurrent articulation and the WLE—A second piece of data that has been taken as evidence for a role of articulatory-based rehearsal in STM is the effect of concurrent articulation. Baddeley et al. (1975; Exp. 7) found that concurrent articulation during list presentation eliminated the WLE in immediate serial recall of visually presented lists. They concluded that it prevented either rehearsal or recoding. To distinguish these possibilities they studied its effect on auditory span, and found that it did not reduce the WLE. This provided evidence that concurrent articulation during list presentation prevented recoding of orthographic to phonological representations, a conclusion that Baddeley et al. endorsed. The results, however, provide no evidence for the role of articulatory rehearsal in preventing decay of phonological representations in the PS.

Baddeley et al. (1975) also noted that, although concurrent articulation did not reduce the WLE in immediate serial recall with auditory presentation, overall performance was reduced under concurrent articulation. They said this could be because articulation may provide “an advantage” to recall performance due to the use of a “supplementary phonemically based store”, which they tentatively link to “an output buffer of some type ... [a] buffer store [that] is necessary for the smooth production of speech”. However, as they note, concurrent articulation may take up “general processing capacity”, and the way to see the results may be as a reduction in recall under concurrent articulation due to demands of a competing task

rather as than as enhancement of recall by an articulatory process in the absence of concurrent articulation. We discuss the general issue of what tasks interfere with immediate serial recall below.

Baddeley, Lewis, and Vallar (1984) extended Baddeley et al.'s (1975) studies of the WLE in auditory immediate serial recall to include suppression during recall by using a written response (participants wrote the long words in abbreviated form—*hippo* for *hippopotamus*—to eliminate differences in output time for lists of short and long words). Concurrent articulation largely or completely abolished the WLE. Baddeley et al. (1984) concluded that, for auditory presentation, “if articulatory coding is to be avoided, it is crucial that subjects be required to suppress articulation throughout both presentation and recall” (p. 245). They focus on the difference between the results for visual lists, where concurrent articulation during presentation eliminates the WLE, saying (p. 245), “this precaution apparently is not required when presentation is visual (Baddeley et al., 1975). The most likely interpretation of this difference would seem to stem from the assumption that articulatory repetition of auditory items is a highly compatible skill that can be performed rapidly and with minimal processing demand.” Hanley and Broadbent (1987) suggested that the finding that the auditory WLE persists under concurrent articulation during presentation only, and disappears under concurrent articulation during both presentation and response, is evidence for articulatory rehearsal during recall.

The suggestion that rehearsal can take place in the presence of concurrent articulation during auditory list presentation requires that rehearsal occur during inter-utterance pauses in concurrent articulation and/or during articulation itself. The problem is that, if circularity is to be avoided in attributing the elimination of the WLE by concurrent articulation to interference with rehearsal, there must be a basis for determining when rehearsal does not occur other than the loss of the WLE, and there is no such independent criterion in this work. Also, the finding cited above that inter-response intervals are not affected by word length argues against the view that the disappearance of the auditory WLE under concurrent articulation during presentation and recall reflects articulatory rehearsal of items at recall.

Concurrent articulation and the PSE—A more direct approach to the locus of the effect of concurrent articulation is to examine its effect on the PSE. The PSE is useful because it is a litmus for a factor affecting the PS. According to Larsen and Baddeley (2003), “the presence of an acoustic similarity effect [i.e., the PSE] will provide prima facie evidence that acoustic coding is involved” (p. 1252). Baddeley et al. (1984) found small or insignificant effects of concurrent articulation on the auditory PSE effect, and interpreted the results as evidence for an automatic entry of spoken words into independent PS that is responsible for the PSE. A problem for this analysis is that the WM model maintains that the PS cannot support spans of six or seven because of decay; if concurrent articulation abolishes articulatory based rehearsal, it is not clear how the PSE arises for longer lists under suppression.

Conclusions concerning the effect of concurrent articulation—Evidence from concurrent articulation supports the view that articulation is needed for recoding of orthographic to phonological representations. However, evidence from concurrent articulation does not provide support for the view that articulatory rehearsal supports items in the PS. As noted above, the overall reduction in performance in immediate serial recall under concurrent articulation could suggest a role for articulation in supporting items in the PS but could also reflect demands of a competing task.

The phonological similarity effect

Unlike the duration-based WLE, the PSE is not in dispute: the advantage of phonologically dissimilar over phonologically similar items in immediate serial recall has been demonstrated repeatedly (Baddeley, 1966a, 1968; Conrad & Hull, 1964; Henson, Norris, Page, & Badeley, 1996) and is considered a major piece of evidence for a role for phonological codes in verbal STM. The questions that arise are where and how the effect arises. Does it occur during encoding, storage or retrieval? What does it imply about the structure and mechanisms of memory?

Locus of the PSE and its implications for the PS—Baddeley (1968) found that recall of phonologically similar items declined less rapidly than that of phonologically dissimilar items in the Brown-Peterson paradigm, indicating the PSE does not arise during storage, and found no effect of presentation in noise on recall or on the PSE, suggesting the PSE does not arise in encoding (entry into STM). He concluded that it arose at retrieval.

The PSE arises for recall of order but not item information (Wickelgren, 1965). Both of the major mechanisms that have been proposed to underlie retrieval of order information—item-to-item associations (“chaining”: Wickelgren, 1965, 1966; Murdock, 1993, 1995), and an “address” mechanism (Crossman, 1959)—can account for the restriction of the PSE to recall of item and order information, but chaining models have difficulty accounting for the sawtooth effect of alternating similar and dissimilar items (Baddeley, 1968; Wickelgren, 1965, 1966). Henson et al. (1996) found that “first-in-report” errors for dissimilar items did not differ in mixed and pure visual lists (Exp. 2) or auditory lists (Exp. 3), and that relative errors did not differ between mixed and pure visual lists (Exp. 1), arguing against a cueing locus for the PSE.

The conclusion that PSE reflects “overloading of retrieval cues which consequently do not discriminate adequately among available responses” (Baddeley, 1968, Abstract, p. 249) is critical. It states that the PSE arises because of the importance of phonological retrieval cues in recall of serial order in short lists at short delays. It does not require that items in STM be maintained only in phonological form, only that phonological information be one of several features of items in an STM store. An effect of phonological similarity at recall is a natural consequence of Luce-type choice rules (Luce, 1959, 1963) in which retrieval is affected by the degree to which the retrieval cue overlaps with features of the target, modified by the degree to which it overlaps with features of competitors. The PSE is therefore consistent with models that postulate a purely phonological store, but it is also consistent with models that do not and instead focus on cues.

Variability of the PSE—Although the PSE is robust, it is not universally present. Baddeley (1966a, 1966b) reported that phonological but not semantic similarity had a great effect on immediate serial recall at list length 5 and that the effect of these factors was reversed at list length 10. The PSE was not found in some poor readers (Mann, Liberman, & Shankweiler, 1980; Shankweiler, Liberman, Mark, Fowler, & Fischer, 1979) but was documented in others (Johnston, 1982), the difference apparently due to the fact that, in studies using fixed list length, the poor readers were tested above span in some studies and at or below span in others (Hall, Wilson, Humphreys, Tinzmann, & Bowyer, 1983; Johnston, Rugg, & Scott, 1987). The auditory PSE persists under concurrent irrelevant speech conditions at list lengths 5–7, and disappears at list length 8 (Salame & Baddeley, 1986).

Baddeley (2000a) suggested that participants abandon the use of phonological coding “when they consistently fail to remember sequences correctly” (p. 544); a “strategy switch”. Strategic abandonment of phonological coding is invoked to account for other phenomena,

and risks circularity. It also encounters several difficulties. One is that it seems to require that the use of the PS be all or none: at longer list lengths participants do not use it at all, and at shorter lengths they do not use the LTM/EB systems. If immediate serial recall depended on both the LTM/EB systems and the PS, one might expect that the PSE would be reduced, or only occur in certain serial positions, at longer list lengths; it is also hard to see how participants turn off a store into which entry is automatic. The mechanism of a “strategy switch” above span also cannot explain the fact that many participants show no PSE or a reverse effect at list lengths at or below their span (9% and 7% in studies by Beaman, Neath, & Surprenant, 2007; Logie, Sala, Laiacona, Chalmers, & Wynn, 1996; and Della Sala & Logie, 1997, on a single test occasion). Beaman et al. (2007) have modelled individual differences in the PSE by varying a single parameter, reflecting attention, that scales the measure of the mismatch between a retrieval cue and an item in secondary memory in a simulated feature-based model of recall that has no memory system devoted to storage of phonological information. Brown, Neath, and Chater (2007) account for much of the PSE data, including individual variability, in a model in which items are represented in terms of their location in multidimensional space, one aspect of which are Weber-compressed temporal intervals between items. These models provide more productive accounts of variability than unconstrained invocation of strategic use of a special store that contains only phonological representations.

Other factors affecting immediate serial recall—The converse to the fact that the PSE is not always seen is the finding that phonological similarity is not the only factor that increases errors in span. Lewandowsky and Oberauer (2008, p. 880) comment that:

... although those two variables [phonological complexity and phonological similarity] are clearly among the most important ones that are known to determine memory, numerous other features of words play a similar role. For example, memory span for words is affected by their familiarity (Hulme, Maughan, & Brown, 1991), their imageability (Bourassa & Besner, 1994), their phonological neighbourhood size and neighbourhood frequency (Allen & Hulme, 2006; Roodenrys, Hulme, Lethbridge, Hinton, & Nimmo, 2002), their concreteness (Walker & Hulme, 1999), and the frequency of their constituent bigrams in the language (Thorn & Frankish, 2005).

There seem to be two theoretical possibilities: either there are multiple stores, or items are represented along multiple dimensions within a single store. The latter is more parsimonious on two grounds: theoretically, it postulates fewer stores; empirically, there already is a store where multiple dimensions are bound: LTM. We note that the former, less parsimonious, possibility diminishes the role of the PS; the latter denies the PS exists.

Conclusions concerning the PSE—We are left with the conclusion there is no clear evidence from the PSE for a PS. The PSE may arise because of the importance of phonological cues to retrieval of item and order information at short intervals when lists are short. This requires an explanation, but does not require that there be a PS.

The irrelevant speech effect (ISE)

Colle and Welsh (1976; see also Colle, 1980) reported that visual digit span was reduced by concurrent unattended (irrelevant) speech. Salame and Baddeley (1982) related the ISE to the WM model. Their conclusion was that irrelevant speech entered the PS and interfered with its function. The effect of unattended speech on span has been investigated in many studies, in part because of its potential importance in the area of effects of environmental noise as well as because it bears on many theoretical issues. We will focus on two sets of

data: (a) the effects of irrelevant speech on the PSE in written and auditory immediate serial recall, and (b) the specificity of the ISE.

The ISE and the PSE—Surprenant et al. (2008) reviewed 12 studies and found that the PSE was reduced by 9% on average from control to unattended (irrelevant) speech conditions (which we found is significant by *t*-test). They argued that “the presence of irrelevant speech either eliminates or greatly reduces the magnitude of the [PSE]” (p. 142). This argues against a “mnemonic masking” model of the ISE that predicts that phonologically similar items, having fewer distinguishing cues, would be more subject to disruption by irrelevant speech (Salamé & Baddeley, 1986).

Larsen and Baddeley (2003) argued that irrelevant speech reduces the PSE by leading participants to abandon phonological codes (a second instance of strategic abandonment of the PS). In support of this analysis they argued that the effect of irrelevant speech on the PSE is list-length dependent, being less marked at shorter lengths (Salamé & Baddeley, 1986). They cited the PSE of 22–23% in the control (no interference) conditions as evidence for the difficulty of the similar sequences. There are serious problems with this argument, however. One is that the proportion of items recalled in correct positions in the similar lists in the control conditions in Larsen and Baddeley (2003) ranged from .54 to .64, far above floor and higher than the (arbitrary) cut-off of 50% that Salamé and Baddeley (1986) and Baddeley (2000a) established as the point at which participants stop using the PS. A second is that Page and Norris (2003) reported unpublished data showing that unattended speech affects recall of phonologically similar items more than phonologically distinct items in mixed lists, which rules out an explanation based on list-wise strategy change. The WM model thus does not adequately address details of the irrelevant speech effect on the PSE.

The specificity of the ISE—A major area of study regarding the ISE is the specificity of the effect. There are two aspects to this issue: does the effect depend on the irrelevant stimulus being speech, and does concurrent unattended speech selectively affect STM and variables in STM that are associated with specific slave systems? The further a concurrent stimulus can be from speech and still interfere with STM and variables that reflect phonological coding in STM, and the less memory-like the tasks that concurrent unattended speech can interfere with, the less relevant the ISE is to models of the memory mechanisms underlying span.

Other interference effects in STM—Several studies have demonstrated considerable similarities between the effects of irrelevant speech and those of non-verbal tasks on STM. Surprenant, Neath, Bireta, and Allbritton (2008) studied irrelevant speech and tapping, and found the same pattern of interaction between condition (no interference, interference) and the list type (phonologically similar, dissimilar) for both tasks and significant correlations between the magnitude of the ISE and the effect of concurrent tapping on immediate serial recall at list length 6. Larsen and Baddeley (2003) studied effects of three interference conditions—irrelevant speech, concurrent articulation, and concurrent tapping—on the PSE in visual immediate serial recall at list length 6, presenting the concurrent tasks as single items that occurred at temporally regular intervals (Exp. 1), as single items that occurred in a syncopated rhythm (Exp. 2), and as series of six items that occurred at temporally regular intervals (Exp. 3). They found that the PSE was not reduced by either regular irrelevant speech or tapping, and that it was reduced equally by these two concurrent tasks in the syncopated and six-item conditions. Tapping affected recall of phonologically different items more than irrelevant speech in the syncopated and six-item conditions and affected recall of phonologically similar items more than irrelevant speech in the syncopated condition. These results indicate that reductions in the PSE and in performances in

immediate serial recall are affected by concurrent tapping as much as, or perhaps more than, by irrelevant speech.¹

These results suggest that the effect of unattended speech on STM is part of a broader picture of effects of concurrent irrelevant stimuli on STM. An attempt to capture the broader picture is found in the work of Macken, Jones, and Hughes, which we will review briefly. These authors developed an “object-oriented episodic record (O-OER)” model of STM to which they relate interference effects (Jones, 1993; Jones & Macken, 1993; Jones, Macken, & Murray, 1993; Jones, Hughes, & Macken, 2006, 2007; Macken & Jones, 1995). In the O-OER model each different item is represented as a different amodal “object”. Serial order is represented by pointers to these objects, which are formed probabilistically and are subject to decay. Different sets of items, including those in different modalities, form different streams. Recall errors occur when pointers from one stream interfere with those in another. The O-OER model accounts for several features of the ISE noted above. For instance, since a repeated object has only one associated pointer, interference will be greater from different than from repeated items. From the present perspective the most important aspect of this model is fact that the features of objects other than their phonological properties determine the ISE, which allows the model to account for effects of other types of concurrent stimuli, not simply sound, on STM.

The fact that the ISE is smaller for unvarying and repeated than for varying stimuli suggests an attention mechanism. In Macken and Jones’ view, unattended stimuli interfere with cognitive activity through two different processes, termed “interference by process” and “attention capture”. Single events lead to attention capture; ongoing unattended stimuli produce “interference by process” through interference of pointers in one stream with those in another. Unattended auditory stimuli cause greater “interference by process” as a function of the degree of acoustic difference between successive irrelevant tokens up to the point of establishing separate streams (Jones, Alford, Bridges, Tremblay, & Macken, 1999; Jones & Macken, 1995a, 1995b; Macken & Jones, 2003; for streaming, see Bregman, 1990) and of the ease of establishing order in the unattended list (Macken & Jones, 2003). Macken, Phelps, and Jones (2009) found that individual differences in the ability to detect changes in auditory sequences in which mismatches could be detected under concurrent articulation correlated with the effect of concurrent lists with this property on serial recall, while individual differences in the ability to detect changes in auditory sequences in which mismatches could not be detected under concurrent articulation did not, leading Jones, Hughes, and Macken (2010) to conclude that “it is the type of auditory processing that may be accomplished in an obligatory, rather than a deliberate, mode of processing that is responsible for interference from changing-state auditory sequences” (p. 206).

¹The effect of concurrent tapping in Larsen and Baddeley (2003) is also relevant to their suggestion that the lack of an effect of irrelevant speech on phonologically similar lists is due to participants abandoning the PS. Concurrent syncopated tapping and articulation produced statistically indistinguishable lower spans in the phonologically similar condition. If the PS was not used when items are phonologically similar, then the reduction in span in the phonologically similar conditions due to concurrent syncopated tapping and concurrent articulation must be because of interference of these concurrent tasks with the CE, not the PS. The claim that concurrent syncopated tapping interferes with the CE is potentially consistent with the WM model (although it is not the view that Larsen & Baddeley, 2003, take), but the idea that concurrent articulation does so is obviously not. A second challenging result of this study is that the effect of concurrent regular single-item tapping was the same as that of irrelevant speech in not reducing the visual PSE. Larsen and Baddeley (2003) say that “the fact that tapping does not reduce the phonological similarity effect is consistent with the assumption that its impact is on the central executive, not on the phonological loop” (p. 1263). By this logic, irrelevant speech might also affect the CE.

The fact that Larson and Baddeley (2003) found no difference in the effect of concurrent tapping and concurrent articulation in Exps 2 and 3 suggests that these interference conditions operate at a common functional locus. Larson and Baddeley suggested that both concurrent syncopated tapping and concurrent articulation disrupted articulatory rehearsal by interfering with a timing process, a suggestion they attribute to Saito (1993, 1994) and that they consider to be a “modification” of the WM model. The idea that concurrent syncopated tapping selectively interferes with rehearsal by interrupting its timing is based on the view that production of articulatory timing is irregular. However, Guérard, Jalbert, Neath, Surprenant, and Bireta (2009) showed that regular tapping of spatially varied targets had the same effect as syncopated tapping, arguing that syncopation is not the critical factor.

Effects of irrelevant speech on other functions—The opposite side of the coin of other interference effects in STM are effects of unattended speech on non-verbal tasks. Farley, Neath, Allbritton, and Surprenant (2007, Exp. 1) found that concurrent unattended speech interfered with performance on a task in which participants identified four visually presented symbols by pressing on corresponding keys. Neath, Guérard, Jalbert, Bireta, and Surprenant (2009) found that unattended speech eliminated (Exp. 1) or reduced (Exp. 2) implicit learning of repeated sequences of nonsense (not verbally codeable) geometric shapes (“statistical learning”; Saffran, Aslin, & Newport, 1996; Turk-Browne, Jungé, & Scholl, 2005). Support for two critical claims of the O-OER—that changing state is critical to the effect of irrelevant speech and that sequence formation is a critical feature of cognitive tasks with which irrelevant speech interferes—comes from details of the effect of unattended speech on these two tasks. Farley et al. (2007, Exp. 1) found that concurrent unattended speech interfered with performance only if the shapes occurred in a pattern, not randomly. Learning, as evidenced by faster RTs in later blocks of trials and slower RTs on a block that contained a different sequence, was not affected. In Exp. 2 they showed that the effect of performance was greater for changing than for repeated unattended speech. Neath et al. (2009) found that the effect of unattended speech on implicit learning occurred only with changing, not repeated, speech. These results are consistent with the changing state hypothesis, which covers the data considered in the WM account of the ISE as well as much additional data.

Conclusions concerning the ISE—The picture that emerges is that the ISE is one of many situations in which there is interference with the representation and/or recall of serial order by items in an unattended stream. Much remains to be learned about the enormous topic of how unattended tasks interfere with various cognitive functions. However, existing work indicates that ISE needs to be considered within a much larger framework and cannot at present clearly be attributed to unattended speech entering the PS.

Serial position effects

The WM model does not model memory for order, but there has been some discussion of issues related to order in the literature on the model. We will discuss two: (1) the fact that the recency effect in free recall does not show sensitivity to the factors that affect immediate serial recall, and (2) an integration of the WM model with models that account for serial position effects.

Factors affecting recency and span—Baddeley et al. (1975) pointed out that span and the recency effect dissociate with respect to the WLE. Craik (1968a) found no WLE in free recall. Watkins and Watkins (1973) found an interaction of word length and serial position in immediate serial recall such that short words were better recalled than long words in the first half of the list and there was a non-significant effect in the opposite direction in the second half (see also Watkins, Watkins, & Crowder, 1974). Glanzer and Razel (1974) found a constant recency effect regardless of length, even for proverbs.² Recency is also not affected by phonological similarity (Craik & Levy, 1970; Glanzer, Koppelaar, & Nelson, 1972) or by articulatory suppression (Richardson & Baddeley, 1975). Baddley and Hitch (1974) found that the recency effect was unaffected by concurrent articulation and almost unaffected by a concurrent immediate serial recall task. Ward (2001) noted that the continuous distractor paradigm (Bjork & Whitten, 1974; Glenberg, 1984; Glenberg et al., 1980; Tzeng, 1973), in which participants perform a distractor task immediately after presentation of each list item, reduces free recall overall but not the recency portion of the free recall curve, inconsistent with recency being supported by a PS subject to overwriting

²An exception is Watkins (1972), who found an equal length effect at all serial positions.

and decay in the absence of rehearsal. Craik (1970) found that span correlated better with the secondary memory than with the primary memory component of free recall.

These findings prompted Baddeley and Hitch (1974) to postulate that recency resulted from the CE and span from the articulatory (later “phonological”) loop (PL). Baddeley et al. (1975) endorsed the Baddeley and Hitch (1974) position that recency and span are supported by different stores. This raises the question of the relation of the CE to span under normal conditions. It is implausible to believe that the CE is preempted by the PL system under normal conditions, so its resources would presumably contribute to span. Similarly, it is implausible (to us, at least) that the PL would not be used at all to derive recency effects. Regardless of how these issues are resolved, the WM model postulates that three stores and processes contribute to span—the CE (CS/EB), the PS, and the articulatory-based rehearsal mechanism—and one, the CE (CS/EB), contributes to recency effects in free recall. Even if span is an emergent property from multiple (memory) systems, this is an unsatisfying state of affairs from a theoretical point of view.

Serial order: Integration of the WM and other models—Recency is one aspect of order effects; the fact that the PSE arises for recall of item and order information but not item information alone is another. These STM phenomena lead to the question of how the WM model accounts for serial order. The simple answer is that it does not attempt to; the WM model as developed by Baddeley does not have a mechanism that has been specifically applied to memory for order. Order information might be available from several features of the WM model, but it is not clear from the presentation of the model which, if any, of these features the advocates of the model believe is responsible for memory for order, or whether any of these features is capable of providing the information needed to retrieve order.

Several researchers have begun to address this issue, integrating some aspects of the WM model with mechanisms that support recall of order. Page and Norris’s (1998, 2003) primacy model, the Burgess and Hitch (1992, 1999) connectionist model, and the start-end model (Henson, 1998) exemplify this work. We will briefly review Page and Norris’s (1998, 2003) model to illustrate one way this connection has been made.

In the primacy model order is coded as a primacy gradient in which items decay with a half life of about 2 seconds. Items in the primacy gradient are not coded solely in terms of phonological features, and their decay is independent of phonological features of items. Item activation levels are increased by cumulative (as opposed to repetitive or associative) rehearsal. Items are forwarded to a second set of nodes, in which their phonological properties are represented, on the basis of their level of activation, with suppression of previously transferred items. Items in the second stage are activated as a function of the product of their activation in the primacy gradient and the phonological similarity of the second stage item to the item in the primacy gradient as determined by a Luce-type rule, both subject to noise. Phonological similarity effects result from confusability with similar items at the second stage. Page and Norris (1998) consider the second stage to be an output stage, and Page and Henson (2001) say that it is motivated by two stage models of word production (Dell 1986; Levelt, 1989). Page and Henson (2001) say, “the primacy model is essentially an implementation of the PL [the phonological loop; i.e., the PS and the articulatory rehearsal mechanism]” (p. 183).

We will not consider the adequacy of the simulations of data on the part of the primacy model here, but rather what features of the model contribute to the recall of order and effects of order on recall. These features involve significant departures from ones found in the WM model.

Order information in the primacy model results from the primacy gradient and rehearsal. However, the primacy gradient does not contain purely phonological representations. This raises several questions. The first concerns rehearsal. Rehearsal in the model is clearly articulatory: Page and Norris (1998) obtained values for the covert rehearsal rate from measurements of reading and repetition rates made by Baddeley et al. (1975), Baddeley and Andrade (1994), and Hulme et al. (1991). However, rehearsal takes place on representations in the first stage of the model, which are not purely phonological, and not on phonological representations in the second set of nodes, a difference from the locus of its operation the WM model. The second concerns the origin of the PSE in a second processing stage associated with speech planning, whereas it is attributed to the PS in the WM model. In this connection it is worth noting that the versions of both Dell's and Levelt's models cited by Page and Henson (2001) deal only with speech production and neither has an input phonological store akin to the PS. Later versions of Dell's model, which do have an input phonological store (e.g., Dell, Schwartz, Martin, Saffran, & Gagnon, 1997), would seem to be closer to the Burgess and Hitch (1999) model, which "cycles" phonological representations through input and output stores.

Thus, although the primacy model includes both a rehearsal mechanism that is closely linked to articulation and a store that represents phonological information, it differs from the WM in two important ways. First, the critical mechanism that supports recall of order information in the primacy model—the primacy gradient—is not found in the WM model. Second, the relation between rehearsal and phonological storage differs in the two models. Similar comments can be made about other models that include both a mechanism that supports recall of order information and a set of phonological representations. For instance, Burgess and Hitch (1992, 1999) include a set of context nodes that are critical to recall of order information, that are not present in the WM model. These models, and others such as Henson (1998), were designed in large part to model memory for order. Many of them include a phonological store and/or an articulatory-based rehearsal mechanism, but we feel it is unreasonable to say that they are "essentially ... implementation[s] of the PL". The discrepancies between the way phonological representations and articulatory-based rehearsal appear in implemented models that account for recall of order information and in the WM model raises questions about whether the WM model contains an architecture of phonological representations and articulatory-based rehearsal that can account for recall of order and serial position effects.

Overview of studies in normal individuals

There is substantial evidence that immediate serial recall and related tasks such as serial reconstruction and free recall are particularly sensitive to phonological features of verbal stimuli, at least when relatively small numbers of items must be remembered. The WM models presents hypotheses about the memory mechanisms that underlie this effect. This review of work since roughly 1990 has at least raised serious questions about, if not undermined, the basic components of the WM model. Specifically we have argued that there is no strong evidence for the existence of a purely phonological store that decays rapidly. Similarly we have argued that there is no evidence that articulatory rehearsal occurs during the presentation, maintenance, or retrieval of items in immediate serial recall. The role of articulation is limited to production of verbal output and recoding of written input. Alternate models, such as the object-oriented episodic record model, extend to other phenomena better than the WM does. Models that account for recall of order information rely on mechanisms not found in the WM model and have functional architectures that relate phonological representations and articulatory rehearsal to one another in different ways than the WM model does.

NEUROPSYCHOLOGICAL CASES

Neurological patients whose STM and related functional capacities have been studied number in the dozens if not hundreds. Our focus is on a selected number of studies, mostly well known, that have been taken as providing important data about the WM model. The discussion is organised by the deficits that have been attributed to the patients: deficits in the PS, deficits in the articulatory rehearsal process, deficits in language functions that may be related to STM, and deficits in other stores that support STM.

PS cases

Shallice and Vallar (1990) argue for the existence of a STM syndrome on neuropsychological grounds. They list 21 “extensively studied STM” cases (their Table 1.1); our discussion will focus on one—PV (Basso, Spinnler, Vallar, & Zanolio, 1982; Vallar & Papagno, 1986, 1995, 2002)—and will mention others at times. Typical patients with the STM syndrome had reduced spans, rapid forgetting in the later trials of the Brown-Peterson task (where proactive interference has plateaued), and a reversed visual-auditory effect in span. Shallice and Vallar (1990) attributed these performances to a deficit in a “phonological buffer store”, which is the equivalent of the PS. They argue that there are two features of the STM cases that support an “input” store view. The first is that the visual-auditory reversal is easily explained if the PS is an input store (and a second visual store can support span) but requires “an additional, rather implausible, assumption” (1990, p. 30) if the PS is an output buffer. They do not state what the additional assumption is, but we infer that it is the two input modalities feed two different output buffers; Howard and Nickels (2005) appealed to this assumption to explain large auditory-visual span dissociations in STM patients showing a visual length effect. The second is that some STM cases did not have speech production deficits (e.g., JB; Shallice & Butterworth 1977).

We will consider this syndrome with respect to the questions of whether the pattern of performance is consistent with the WM theory or extends or modifies the theory in natural ways. We will focus on PV, one of the best-studied cases, mentioning other cases at times.

The locus of the deficit in PS cases—The first issue we will consider in examining the first question is what the deficit is in these cases. Shallice and Vallar, and other researchers (e.g., Baddeley, 1998) have argued that it lies in the PS. Two challenges to this view that Shallice and Vallar identify are that PV did not show a WLE, which would have occurred had she rehearsed, and that, where tested (KF and PV), the “PS” patients retained the auditory PSE.

Shallice and Vallar account for the absence of the WLE in PV by saying that PV made a strategic choice not to rehearse, because of her defective PS and the relative lack of utility of rehearsal. This is another appeal to strategic abandonment of a mechanism that supports STM performance, and subject to the concerns raised above regarding the abandonment of the PS. We also think the abandonment hypothesis is unconvincing on these grounds in patients as well as in controls; our view is that it would be just as rational for PV to utilise as many means as possible to accomplish a task, given a defective component of the system, as to abandon a less-than-maximally-effective means of enhancing performance.

In our view the more serious problem is presence of the auditory PSE (see Caplan & Waters, 1990, for comments). Shallice and Vallar account for the retained PSE by arguing that, if phonologically confusable sequences require more capacity than non-confusable sequences, a pathologically limited capacity in the PS will reduce span without abolishing the PSE. As we have seen, this view conflicts with that of other researchers who have argued that the

PSE is evidence for the integrity of the PS, and that reduction in the functioning of the PS (e.g., due to irrelevant speech) is expected to eliminate or reduce the PSE.

In coming to their conclusion, Shallice and Vallar reference Sperling and Speelman (1970) for evidence that phonologically confusable sequences require more capacity than non-confusable sequences. Since this is one of the few discussions of the relation between STM capacity and the PSE, we shall review the discussion. Sperling and Speelman (1970) presented a “phonemic” model of auditory STM in which memory is for phonemes only. They calculated the letter capacity for lists of dissimilar letters and the decrement in capacity for lists of similar letters, and found that the capacity deficit for similar item lists was 2.22, 2.21, and 2.31 for presentation rates of 1/second, 2/second, and 4/second, respectively. Shallice and Vallar (1990) are therefore correct in their statement that, in Sperling and Speelman’s model, there is less capacity for phonologically similar than for phonologically dissimilar items. However, this does not imply that the PSE survives capacity reduction; Sperling and Speelman did not report predictions of the model for the PSE at different levels of capacity. We applied the principles of their model to recall of letters in lists of two and three similar and dissimilar bi-phonemic, mono-syllabic letters at phoneme capacities of 2 and 3, chosen to exemplify situations in which phoneme capacity was less than, equalled, or exceeded the number of letters to be recalled. There were different effects of phoneme capacity at different list lengths. The PSE was larger at list length 2 than 3 (where it was reversed), did not change with phoneme capacity at list length 2, and was smaller with greater phoneme capacity at list length 3. These incomplete applications indicate that the relation between capacity, list length, and the PSE in Sperling and Speelman’s model is complex and that it is not obvious that their model accounts for the PSE in PV or other cases with reduced PS capacity. We also note that Sperling and Speelman’s model is not above criticism; the fact that memory is only for phonemes and not for letters, for instance, is a simplification.

There are also empirical data regarding the magnitude of the PSE as a function of recall of dissimilar items. Sperling and Speelman (1970, Table 2, p. 168), reported the PSE as a function of recall of dissimilar items at list lengths 4, 6, and 8, presented at rates of 1, 2, and 4 items/second. The PSE was considerably reduced at lower list lengths, regardless of presentation rate and regardless of the interaction of presentation rate and list length. One of the authors (DC) has collected data on word and digit immediate serial recall and the auditory PSE in 51 aphasic patients (unpublished data). There was a significant correlation between the PSE and total item and order recall ($r = .27, p = .05$). Again, these observations do not show that the preservation of the PSE in PV is inconsistent with a reduced PS, but they indicate that, contrary to the findings in PV, lower list lengths and reduced spans are associated with a smaller PSE.

The effect of phonological similarity on auditory span has been studied in a number of other patients in whom a PS deficit has been suggested. Most show the auditory PSE (see Caplan & Waters, 1990), but not all, and comparing cases illustrates the challenge in relating the magnitude of the PS deficit to the auditory PSE. For instance, HB (Howard & Nickels, 2005), with an auditory span of 4, did not show a reliable auditory PSE, while PV, with a span of at most 2, did.

The conclusion we draw from these considerations is that PV’s retained PSE constitutes a problem for the view that she has a deficit in the PS. Sperling and Speelman’s model may or may not be able to account for the pattern of preservation of the auditory PSE in “PS” cases; empirically, if we are to extrapolate from performance at different list lengths reported by Sperling and Speelman, the auditory PSE is too large in PV and, we believe, in many other “PS” cases. Looked at the other way around, if the analysis of PV as having a PS deficit is

correct, the role of the PS in generating the PSE is unclear—which raises a major problem for the WM model.

The recency effect—The second problem with attributing a PS deficit to PV is the absence of a recency effect. KF (Shallice & Warrington, 1970) and JB and WH (Warrington, Logue, & Pratt, 1971) have also shown reduced recency effects. As discussed above, factors that affect span (the PSE and WLE) do not affect the recency portion of the free recall curve, leading to Baddeley's attributing recency to the EB not the PS or rehearsal. Vallar and his colleagues, however, attribute the loss of recency in PV to her reduced PS. Is this an instance of a productive application of neuropsychological data to the development of the WM model, or a contradiction?

Shallice and Vallar (1990) deal with the discrepancy between the neuropsychological data and the WM theory based on normal data by arguing that the evidence from normal participants that recency is not supported by the PS has been over-interpreted. They point out that the PSE arises for item-and-order, not item information and thus would not be expected to occur in free recall, and that evidence for phonological representations in recency positions comes from the finding that errors in free recall are phonologically related to correct items only in list terminal positions (Fraik, 1968b; Shallice, 1975). However, Shallice (1975) himself "strongly criticized" (p. 274) Fraik (1968a) in part on the grounds that the definition of phonologically similar error was "loose", which would have overestimated the extent of phonological errors.³ Shallice (1975) had participants recall from end and found that both "acoustic errors" and morphological errors were more common in the last few positions of the recall curve for 15 item lists. The second of these errors types suggest that recently presented items are coded in terms of more than phonological features.

Shallice and Vallar (1990) were mostly concerned that the finding of recency effects in LTM had led to alternate accounts of the effect in STM, such as retrieval strategy (Baddeley & Hitch, 1977, 1993), a form of temporal discrimination (Glenberg et al., 1980), competitive activation (Cowan, 1993; Hinton & Plaut, 1987), competitive queuing (Burgess & Hitch, 1992), and others. Attribution of the absent recency effect in PV to such mechanisms would imply that she has two deficits, one leading to the reduction in span and the other to the absence of recency.

Shallice and Vallar (1990) argued against a temporal distinctiveness account of the absence of recency in PV because her performance was in the normal range with visual lists. We question this characterisation of PV's free recall performance. Figure 1 shows the serial position curves for free recall and recall-from-end conditions for controls and PV from Vallar and Papagno (1986). It is far from apparent that PV has a recency effect in visual free recall, and a difference between her visual and auditory recall-from-end performance occurred only in positions 10 and 11.

To the extent that recency is better preserved in visual than auditory lists in PV, it may be because of modality effects on the use of temporal information as a retrieval cue (see Greene, 1986, for discussion). O'Connor and Hermelin (1972) found that, when three successive digits were presented either visually or auditorily at three different locations such that the second digit never occurred in the central location, children reported the visually presented item from the central spatial location and the auditorily presented item that

³Shallice (1975) also criticised Fraik (1968a) on the grounds that phonological errors could have been perceptual errors and that a fast presentation rate was used leading to reliance on semantic processing, both of which would have underestimated phonological errors. The lesson to be drawn from Fraik (1968a) is thus not clear.

occurred second in the temporal sequence as the middle item. Thus failure to use temporal distinctiveness as a retrieval cue may affect recall for auditory lists more than for visual lists.

Shallice and Vallar argued against a retrieval strategy account of the absence of recency in PV, because of the persistence of this finding in recall from end, despite PV's having an output order comparable to that of normal participants. Although such a strategy, if it was used, could have resulted from a damaged PS, it could have other origins as well. However, there is no clear evidence that PV used such a strategy. To us, the most striking feature in her recall is not a reduced recency but an enhanced primacy effect. PV recalled far more items from the initial position of the list than controls, even when she recalled these items last. Vallar and Papagno (1986) argue that primacy effects in PV's performances reflect the fact that she recalls early list items first in auditory free recall, which they said was a strategic choice regarding order of report occasioned by her deteriorated PS. However, this account cannot explain her better-than-normal pre-recency performances in the recall-from-end conditions, where both she and controls report first items last; these require that some factor enhance recall of early list items, not just that some factor (such as a reduced PS) reduce recall of late list items. PV might report early list items first because her secondary memory is excellent (perhaps better than normal), which might lead to loss of recency; i.e., the cause of her recall strategy could be a positive impetus to maximise a strength rather than to circumvent a weakness.

Finally, other patients have shown different patterns. Della Sala, Logie, Trivelli, Cubelli, and Marchetti (1998) reported two cases, one (AN) with reduced span and retained recency and the other (GC) with intact span and reduced recency. AN showed small PSEs in both visual and auditory span and a WLE in visual, but not auditory span; GC showed normal PSEs and WLEs in span in both modalities. By standard criteria this double dissociation indicates span and recency are supported by different mechanisms. Some caution must be exercised in drawing this conclusion because GC's LTM was impaired, perhaps pointing to abnormalities that affected recency in immediate free recall.

Conclusions concerning PS cases—The deficits seen in neurological cases considered to have PS impairments do not fit the predictions made by the WM model. One problem is the persistence of the auditory PSE in many cases, which does not follow from the theory and is inconsistent with predictions that have been made about the effect of reductions in the PS. The reduced recency effect in some "PS" cases might unite models of span and recency in free recall and thus modify one aspect of the WM model. However, where studied, as in PV, recall performances often show features that are not well understood, so the implications of the data are not clear. Moreover, the conclusion that the PS supports both immediate serial recall and recency is based on the co-occurrence of an alleged PS deficit and reduced recency, and these phenomena have been shown to dissociate.

Cases with articulatory-based rehearsal deficits

Two issues have been discussed regarding articulation in STM: the nature of the articulatory process that is involved in STM, and the nature of the STM-related operations in which an articulatory process is involved.

The nature of articulatory based rehearsal—If the view that an articulatory process supports STM is correct, we would expect that patients with articulatory disorders would have abnormal STM performances; specifically, their performances should look like those of normal participants under concurrent articulation. Reports that some anarthric patients had normal STM performances (Baddeley & Wilson, 1985; Vallar & Cappa, 1987) led to a

revision of the theory: articulation per se was not considered to be required for “articulatory based” rehearsal or recoding; rather, mental preparation for articulation was thought to underlie the WLE and the speech rate-span correlation, and to be disrupted by concurrent articulation. (We will continue to use the term “articulatory-based rehearsal” to refer to these more abstract processes.)

Evidence that mental preparation for articulation underlies the WLE came from the finding that patients with more complex speech output disorders that affected planning of articulatory gestures did show patterns of STM performance that could be attributed to disruption of the articulatory processes. Rochon, Caplan, and Waters (1991) and Waters, Rochon, and Caplan (1992) reported that, as a group, six patients with apraxia of speech showed reduced spans and reduced visual WLE and PSE and auditory WLE, the pattern expected if this condition affects recoding and rehearsal. Waters et al. suggested that the entire picture indicated that recoding and rehearsal require formulation of articulatory plans, but not articulation itself. A problem with this conclusion is that none of the individual patients in Rochon et al. (1991) or Waters et al. (1992) showed the pattern seen on average in the group, raising the concerns voiced by Caramazza (1986) regarding individual performances, the true data from neuropsychology (Shallice, 1988), being “lost in the mean”. Rochon et al. and Waters et al. argued that the loss of the auditory and visual WLE and PSE in their patients followed a comprehensible pattern, with patients with greater impairment (as measured by speech rate) losing more effects, and the loss following an orderly pattern: visual WLE lost with least impairment; combinations of visual WLE and PSE and auditory WLE lost with greater impairment; all effects lost with greatest impairment. The logic of this progression is based on the view that rehearsal and recoding are both needed for the visual WLE, recoding for the visual PSE, and rehearsal for the auditory WLE. This assumes that the demands made by rehearsal and recoding on articulatory-related processes are equivalent, a questionable assumption (see below). The absence of any patient in this series with the expected pattern of WLE and PSE remains troublesome for the conclusion that interruption of planning articulatory gestures produces the pattern of effects that the WM theory associated with articulatory-based rehearsal deficits.

Moving further proximally in the speech production pathway, it would seem to be necessary that patients with speech planning disorders affecting selection and ordering of phonemes but with few or no problems with articulation or planning of actual articulation have articulatory-based rehearsal impairments. One view of articulatory-based rehearsal is as “a working memory space in which phonological segments are temporarily stored prior to the application of various output processes, such as planning and editing of the articulatory procedures needed to produce speech” (Vallar, Di Betta, & Silveri, 1997, p. 795). An output buffer with these properties corresponds to the “phonological output buffer” (POB), a post-lexical “buffer” that many models of speech production posit (Dell & O’Seaghdha, 1991; Dell, Martin, & Schwartz, 2007; Dell et al., 1997; Miceli, Silveri, & Caramazza, 1986), which transmits an ordered set of phonemes to the articulatory planning process; some models integrate the POB with the “phonological output lexicon” (POL). As far as we know, no models of speech production postulate direct connections (i.e., ones bypassing the POB) from any input store (the “phonological input buffer”, the “orthographic input buffer”), the “phonological input lexicon”, the “visual input lexicon”) or semantics to the mechanism that plans articulatory gestures, and therefore these buffers/stores cannot be the source of representations that are used in articulatory-based rehearsal. It would therefore appear that rehearsal could not occur in the presence of significant disturbances of the POB. The Waters et al. (1992) view that rehearsal involves articulatory planning assumed that it required the POB.

It is therefore potentially problematic that some patients with predominantly POB (POB/POL) deficits, called “reproduction conduction aphasics” by Shallice and Warrington (1977), have shown relatively good STM performances. For instance, RL (Caplan, Vanier, & Baker, 1986) made numerous phonemic paraphasias in word and nonword repetition, word reading, and picture naming, yet had a pointing span of five words and was above chance in recognising item order in lists of five words. It is unclear whether this degree of preservation of immediate serial recall is consistent with the degree of RL’s disorder of phoneme choice and ordering given the role of rehearsal in the WM model. However, many details of STM performance are missing (RL was not tested for the WLE or PSE, for instance) and whether this and other “reproduction conduction aphasia” cases (e.g., Damasio & Damasio, 1980) constitute serious problems for the WM model is unclear at this time.

Mechanisms of recoding and rehearsal—The WM model maintains that articulation-related operations support both recoding and rehearsal. We have questioned the evidence for articulation-based rehearsal. The evidence for a role for articulation-based recoding is the elimination of the visual WLE and PSE by concurrent articulation during list presentation. However, the preservation of written homophone judgements under concurrent articulation (Besner, 1987; Tree, Longmore, & Besner, 2011) indicates either that recoding is possible without articulation-based processing or that concurrent articulation does not eliminate articulation-based processing.

Shallice and Vallar (1990) and Vallar et al. (1997) consider a different view of the role of articulatory-based processing in STM and related functions that addresses these issues. In their view, articulatory-based rehearsal has the function of “the recirculation or transfer of phonological information between the PhSTS and a phonological output buffer (or assembly system) involved in the programming of speech output ... The rehearsal process, therefore, may involve a translation between input ‘auditory’ and output ‘articulatory’ phonological codes” (Vallar et al., 1997, p. 795). On this view, concurrent articulation does not prevent recoding (print to sound conversion); rather it prevents the circulation of phonological representations to the PS. This is needed for tasks that make demands on memory—span and rhyme judgement—but not for written homophone judgements, which are preserved under concurrent articulation.

PV’s performances led to an elaboration of this model. PV had intact recoding (shown by her good nonword reading), which led to the formulation of the role of articulatory based rehearsal above, and Vallar and his colleagues argued that she did not rehearse in span tasks for strategic reasons. However, she showed good picture–nonword rhyming judgements which, on the model above, make use of the same articulatory-based, rehearsal-based process that supports rehearsal in immediate serial recall. Invoking strategic non-use of articulatory-based rehearsal in immediate serial recall but not in rhyme judgements is post hoc and ad hoc (and strains credibility). Shallice and Vallar (1990) thus added the hypothesis that immediate serial recall requires more rehearsal than written rhyme judgement or reading nonwords. In this case PV’s performance might be accounted for by a partial deficit in this articulatory-based process.

Shallice and Vallar (1990) also rejected the “unitary” view that articulatory-based recoding and rehearsal utilise the same mechanism because of results in another case, MDC. MDC (Vallar & Cappa, 1987) showed retention of the WLE and PSE in auditory span, loss of these effects in visual span, and defective visual rhyme judgement. MDC was said to have a disturbance of recoding, which precluded the application of articulatory-based rehearsal to visually presented materials. Shallice and Vallar (1990) contrasted the pattern of loss of the visual and retention of the auditory WLE and PSE in MDC with the disruption of the auditory WLE by concurrent articulation in normal participants. Assuming that the WLE,

the written PSE, and written homophone judgement involve articulation-based processes that are ineffective under concurrent articulation and in MDC, these data constitute a double dissociation that points to different articulatory mechanisms being involved in recoding and rehearsal.⁴

Shallice and Vallar thus combine the “two articulatory processes” view with the idea that immediate serial recall requires more use of articulatory based rehearsal than written rhyme judgement. It is not clear, however, how much this would mitigate concerns about how to account for PV’s performances. The claim is that PV’s PS deficit leads to strategic non-use of articulatory-based rehearsal only when the task is more demanding. Why this would lead to its non-use for list lengths of two and not to its non-use for rhyme judgements is not clear, however. List lengths of two are within PV’s span and require retention of the same number of items as rhyme judgements. One might invoke the extent to which the stimuli have to be processed as the basis for saying that one task is harder than the other, but this is hard to assess. Immediate serial recall requires retention of order, not needed in picture–nonword rhyme judgement, and picture–nonword rhyme judgement allows for refreshing one phonological representation from sensory input, but picture–nonword rhyme judgement requires segmentation and comparison operations that are not needed in immediate serial recall and that are known to be difficult.

Conclusions concerning cases with articulatory disturbances—Overall, “articulatory-based rehearsal” cases show that articulation per se is not needed for the functions that the WM model attributes to articulatory-based rehearsal. This “bumps the operations involved in articulatory based rehearsal upstairs” to planning rather than producing speech. However, neuropsychological cases introduce many challenges to accounts of the role of articulation in rehearsal and recoding that are proposed in the WM model.

Cases with other language deficits

The studies reviewed in this section argue that aspects of STM performance are a consequence of word processing, thereby either eliminating the need for specialised stores or modifying views of how these stores operate. We focus on a series of studies by N. Martin and her colleagues; see also studies by Howard and Franklin (1988, 1990), R. Martin, Lesch, and Bartha (1999) and others.

Lexical processing deficits and serial position effects—Saffran and Martin (1990) reported effects of lexical frequency and imageability as a function of serial position in two patients. One (CN) showed reduced recency in digit span but a tendency to recall late list items early in word span, resulting in almost no items reported in correct serial position and progressively more items recalled across positions when order was disregarded. On a test in which she recalled the item in a serial position designated by pointing to a card in a series of items, CN tended to place late list items early, indicating that her span performance was not simply due to a report-from-end strategy. She recalled only two-item nonword lists, and almost only the last position in four-item nonword lists. CN showed an imageability effect in recall of the first list item and trend towards a frequency effect in the last list position. She recalled more early-list items in lists of semantically similar words. CN’s errors were mostly intrusions from earlier lists (not semantically or phonologically related errors) and her recall improved considerably for repeated lists, especially for early positions. The second case, TI,

⁴The double dissociation is unusual in that it involves normal participants under concurrent articulation and a patient, but its implications, such as they are (see Shallice, 1988, for reservations about the implications of double dissociations), are the same as those of the more usually cited double dissociations between patients or across tasks in normal participants (as in Shah & Miyake, 1996).

showed reduced recency effects in digit and word span, effects of imageability in list-initial position and of frequency in the last two positions, and phonological errors.

Saffran and Martin emphasised the effects of lexical and semantic variables (frequency and imageability) on performance. In their view phonological representations activate lexical items, which in turn support maintenance of representations in STM; imageability exerts its effect by reinforcing lexical representations. They argue that an extension of Dell and O'Seaghdha's (1991) model of single word production to list recall predicts semantic and lexical effects to occur in early list positions in normal participants, because the effect of these factors is exerted over a longer time frame than that of phonological factors due to the order in which phonological, lexical, and semantic representations are activated from spoken stimuli and the functional architecture of their interactions. The Dell and O'Seaghdha model interacts with patients' deficits: Saffran and Martin suggest that, in patients with reduced phonological memory functions, lexical and semantic support is more important in late-list positions. This explains the locus of the frequency effect in late-list positions in CN and TI; the weaker frequency effect in CN reflects less lexical support in her than in TI.

The chief problem that Saffran and Martin face is the location of the imageability effect. As outlined above, their analysis predicts that it will arise on late-list positions in patients with phonological problems; the location of the imageability effect on the list-initial item in these patients is therefore a problem. Saffran and Martin attribute this locus of the imageability effect to attention factors, but if imageability exerts its effect through reinforcement of lexical items, an effect of attention that leads to an effect of imageability on list-initial items should also lead to an effect of frequency on those items. Other problems are that the strength of late-list items in CN is unexplained, and the claim that she receives less lexical support than TI is unattested as TI was not tested on lexical tasks, and CN's good performance on abstract words in lexical decision and tachistoscopic reading does not support the view that abstract concepts would only weakly reinforce lexical items.

Martin and Saffran (1997) pursued the effects of semantic and lexical factors on items recalled in different list positions in a series of 15 aphasic patients. Semantic abilities and phonological abilities were assessed, producing "composite S(ematic) and P(honological) indices", and patients were tested on single word repetition and immediate serial recall for two word lists (words in isolation and in the lists varied in frequency and imageability). Martin and Saffran predicted that disorders of phonological processing would lead to increased reliance on lexical and semantic representations in STM, leading to increased effects of imageability and frequency (negative correlations of these measures with the composite P index), and semantic impairments would be expected to show the reverse pattern. Martin and Saffran also made predictions regarding the serial positions in which effects of semantic and phonological deficits should be seen. The argument regarding position effects in lists is based on the view that semantic variables affect early list positions for the reasons outlined above; the argument regarding position effects in words is based on the observation that resolution of the TOT state is associated with access to word-initial phonology (Kohn et al., 1987) and with faulty word retrieval in aphasia (Goodglass et al., 1997). On this basis, patients with better semantic processing should show better performance in first positions in both lists and words, and patients with better phonological processing should show better performance in last positions in both lists and words.

There were positive correlations of the composite S index with primacy effects in lists and words, and of the composite P index with recency effects in words; the correlations of the composite P index with primacy effects in words were negative.⁵ Primacy effects in lists and words were positively correlated with imageability effects in single words and the second word in lists; the primacy effect in words was positively correlated with the frequency effect

in the first word in lists. Martin and Saffran (1997) argue that these results support an interactive, multilevel, lexical-processing-related model of STM. We see several problems with this conclusion.

The first are empirical issues. The correlation of imageability and production of the second word in the two-word lists in this study contrasts with the locus of the imageability effect on the first list item in CN. Other inconsistent data come from semantic dementia patients, some of whom have shown preservation of list-initial items in list recall (Knott, Patterson, & Hodges, 1997). Since these patients have less semantic support for lexical items, primacy effects should be reduced. Martin and Gupta (2004) argue that the difference between the semantic dementia and vascular cases is due to semantic dementia patients losing semantic knowledge and vascular aphasic patients losing processing abilities, but it is hard to see how processing of severely degraded semantic or conceptual representations could effectively support lexical representations. Yet more inconsistencies come from Martin and Gupta's (2004) report that they found both frequency and imageability effects in immediate serial repetition in 19 aphasic patients, but no serial position effect of these variables. Martin and Gupta argue that no serial position effect of these factors would be expected, a conclusion at odds with the Martin and Saffran (1997) model, as best we can see.

From the modelling viewpoint, there is no mechanism in Dell and O'Seaghdha's (1991) model that produces semantic or lexical effects on phoneme production in different positions within words. In both the 1992 and subsequent versions of the model phonological segments within words are activated in parallel from word nodes, and it is not clear how semantic or lexical information affects activation of phonemes at the output planning level at which serial order is utilised.

Martin and Ayala (2004) approached the issue of the effect of language-processing abilities on immediate serial recall by considering a different aspect of performance. They correlated performance on phonological and semantic tasks and phonological and semantic errors in picture naming with item recall in spoken and pointing digit and word immediate serial recall in 47 aphasic patients. Patients were divided into those with relatively intact semantic or phonological processing on the basis of the ratio of their composite S index to the sum of their composite S and P indices. Martin and Ayala predicted that phonological measures would correlate with performance more in the repetition tasks, which require phonology, particularly in patients with more severe phonological deficits, and that semantic measures would correlate with performance more in the pointing tasks, which require semantics, particularly in patients with more severe semantic deficits. They also predicted that semantic measures would correlate with performance more in the repetition tasks in patients with more severe semantic deficits, to the extent that that task involved semantics.

The results were that phonological indices correlated with all recall measures and semantic indices correlated with pointing recall. Phonological error rates in naming correlated negatively with all recall performances and semantic error rates correlated negatively with pointing performances. Patients with better semantic than phonological processing showed no relation of semantic processing with recall, positive correlations between all phonological processing measures (except rhyme judgement) and recall, and a negative correlation of phonological errors in naming with recall. Patients with better phonological than semantic processing showed significant correlations between all semantic and phonological

⁵Martin and Saffran also correlated the difference between the composite S and P indices (the "P-S Index") with these measures, but these correlations with difference scores are hard to interpret (patients with equal P-S Index scores could have very different levels of performance). Correlations involving difference scores are also suspect because difference scores are less reliable than individual scores. The decrease in reliability increases with the correlation of the tests involved in the difference; Martin and Saffran's P and S indices were not correlated, making this issue less important but not eliminating the problem of reliability.

processing measures and recall, significantly negative correlations of semantic errors in naming with all recall performances, and significantly negative correlations of phonological errors in naming with repetition performances.⁶

Martin and Ayala say the results partially support their predictions. They argue that the more frequent correlations of semantic functions with pointing word recall indicates that it relies more on semantic processing than repetition span. Looked at from the point of view of the effect of language impairments on STM, they concluded that semantic deficits affect pointing recall and phonological deficits affect both pointing and repetition recall. The main focus of the discussion was on the effects within the two patient groups. As predicted, semantic abilities and semantic errors correlated with both pointing and repetition recall performances in patients with relatively more impaired semantic abilities but not in patients with more severe phonological deficits. However, contrary to expectations, phonological abilities also correlated with all recall performances in patients with more impaired semantic abilities. Martin and Ayala accounted for these correlations of phonological abilities and recall in patients with relatively more impaired semantic abilities by suggesting that all patients may have some phonological deficits, and that the activation of phonological representations might not be “enduring” in the presence of semantic deficits (which reduce semantic feedback to lexical and thus phonological representations). They argued that the implication of the observed patterns is that the mechanisms of semantic, lexical and phonological processing of words can explain the effects of semantic and phonological deficits on patients’ STM performance, without the necessity of postulating separate phonologically based STM and semantically based LTM systems (i.e., PM and SM). This in turn, they said, argues for an interactive, multilevel, lexical-processing-related model of STM.

There are several issues that arise about these results and conclusions. A methodological issue is the use of item recall as the measure of STM in a task that requires both item and order recall. This is of concern because it underestimates item recall and overestimates span, and may do so differently in different patients. An interpretive issue is that Martin and Ayala argue that the lack of an overall correlation of semantic abilities and repetition performances reflects the ability to perform the repetition span phonologically, without recourse to semantics. This leaves the similar lack of an overall correlation of semantic abilities and pointing digit performances unexplained. The main issue, however, is the results in the subgroups.

These results are more complex than is conveyed in the paper because the division of patients into groups did not result in groups with clear cut differences in semantic capacities. Examination of the data in Martin and Ayala’s Appendix D shows that, although both the Composite P and Composite S indices differed between the ($S > P$) and ($S \leq P$) patients (by t test), the Composite S indices overlapped considerably in the two groups: the Composite S index for 20 of the 23 ($S \leq P$) patients fell within the range of Composite S indices for the ($S > P$) patients, and the Composite S index for 20 of the 23 ($S > P$) patients fell within the range of Composite S indices for the ($S \leq P$) patients. When the three ($S \leq P$) patients with very low Composite S indices and the three ($S > P$) patients with very high Composite S indices (i.e., those whose indices are outside the range of the other group) are removed, the Composite S indices in the groups do not differ, there are no significant correlations between the Composite S index and any recall measure in any group, and the only significant correlations with the Composite P index with recall are for word recall performances in the ($S \leq P$) patients. The different relations of semantic performances in the groups and recall

⁶Martin and Ayala’s (2004) Table 7 (p. 475) appears to contain an error, indicating that there was a correlation between semantic abilities and repetition “modified spans” in patients with more severe phonological deficits ($S > P$ patients).

are thus due to a small number of patients with very good or very poor semantic abilities relative to the group as a whole, and the relation of phonological performances in the groups and recall arises in patients with better phonological skills once patients at the extremes of performance are removed.

The picture is yet more complex. Although the Composite S and P indices did not correlate over all the patients ($r = .14$), inspection of their Table 2 shows that there were strong similarities in the phonological and semantic performances of the six participants whose S indices fell outside the range of the other group's S indices. The three ($S < P$) patients with very low Composite S indices also had low Composite P indices (these included the two lowest Composite P indices in the entire aphasic group) and one of the three ($S > P$) patients with very high Composite S indices had the highest Composite P index in the entire aphasic group. It is possible that there are complex interactions between semantic and phonological processing that affect STM, with patients at the extreme ends of the range of aphasic performance contributing in disproportionate ways to the effects of these abilities on modified immediate serial recall.

In summary, these studies present the view that the effects of semantic and lexical factors differ in different serial positions because of how words are processed in the normal language-processing system; these lexical processing factors interact with language deficits to produce the patterns seen in patients. However, the data from patients consist of aspects of performance that have not generally been modelled in normal participants (item recall in immediate serial recall tasks) and the data are complex and apparently contradictory across studies, requiring unverified assumptions about mechanisms that produce some results (e.g., the role of attention in CN). Although the existence of lexical and semantic effects in these STM tasks in these cases poses a challenge to the WM model, it is hard to judge the success of this particular model in accounting for these effects in either normal individuals or patients with language deficits.

Longitudinal studies—N. Martin and colleagues' second line of research deals with changes in STM performances as a function of changes in lexical processing over time. N. Martin et al. (Martin, Dell, & Schwartz, 1996; Martin, Saffran, Dell, & Schwartz, 1994) studied NC, a patient with deep dysphasia who showed fewer semantic and more phonological paraphasias in repetition over a 2-year period. N. Martin et al. (1994) modelled NC's repetition performance as a pathologically increased decay rate in a version of Dell and O'Seaghdha's (1991) naming model that was modified to correspond to repetition (input consisted of phonological rather than semantic representations); the changes in the proportion of error types over time was simulated by returning the decay rate towards normal.⁷ Martin et al. (1996) added a temporal decay factor to the modified model. They simulated the effect of delay and lists on performance in NC by varying the time between item presentation and selection. Simulation of single word repetition used 8 time steps before selection of word and phonological nodes; simulation of production of the first word of a pair used 9 time steps; simulation of production of the second word of a pair used 10 time steps. Imageability was simulated by varying the number of semantic features that overlapped between words. The effect of delayed production was the same as that of a higher decay rate: increasing the time between presentation and response selection led to fewer errors, more semantic paraphasias, and fewer formal phonological paraphasias. This led to the prediction that, in his (partially) recovered state, NC would show the same pattern of error types when producing items after a delay as he had in immediate item repetition.

⁷Nickels and Howard (1995) note that adapting the Dell and O'Seaghdha model to input (and thus also repetition) is problematic because the adaptation generates 37% phonological errors in comprehension (e.g., "cat" misunderstood as MAT).

before he had recovered. This would be expected to affect immediate serial recall, and particular positions within immediate serial recall.

Martin et al. (1996) say the simulation corresponded to NC's performance. The key results were that, as NC improved, he made fewer errors and his semantic errors disappeared in single word repetition and reappeared in immediate serial recall, particularly in production of items in list final position, for which presentation-to-report time was longer than for list-initial position. Martin et al. took this result as confirmation of their modelling of changes in NC's repetition performance over time by changes in the decay parameter in the lexical processing model over time, and for the claim that the features of the lexical processing model account for some of the features of NC's STM performance.

We question how close the simulation is to NC's performance. Figure 2 shows the results of CN's performances taken from Martin et al. (1996, Figs 5 and 6) in terms of the number of semantic, formal lexical and neologistic paraphasias in three conditions—production of the first and second words of two word lists and repetition of single words after unfilled delays (descending left-sided panels)—for high-imageable (left) and low-imageable (right) words, and plots of the corresponding response types extracted from their Tables 2 and 4 for high-imageable (left) and low-imageable (right) words in simulations with 9 and 10 time steps.⁸ The simulation using 9 time steps between stimulus and recall (top pair of panels on the right) corresponds to performance on the first word in a list (top pair of panels on the left). The simulation using 10 time steps correspond to performance on the second word in a list (middle pair of panels on the left). Both simulations might be considered to correspond to performance after an unfilled delay (bottom pair of panels on the left). Martin et al.'s depiction of the number of correct responses and semantic errors in the simulations averaged over time points (their figure 7, p. 104) is also displayed. The increase in semantic errors in high-imageable words shown in their figure is due to the increase in such errors at the longer delay (10 time steps). The striking feature of these figures to us is that, aside from this increase in semantic errors, the simulations correspond poorly to NC's error patterns. NC makes many more neologisms in the first word of lists than appear in the model with 9 steps, and far fewer formal errors than the model with either 9 or 10 time steps. Even in the one area where the model partially simulates NC's performance—the increase in semantic errors in the second position in lists—there are differences: NC makes many more such errors than the model. An issue that needs to be considered is how to assess the adequacy of the model. Martin et al. (1996) do not provide measures of fit, such as R^2 , AIC, or BIC. There are no criteria for accepting a model based on visual comparison of a single test statistic to a single result in a model. However, it appears that some additional lesion needs to be made in the model to simulate the entire pattern of results. Martin et al. (1996) discuss other possible aspects of the model that could be lesioned; in particular, they suggest that the decay rates and weights between semantic and lexical nodes and between lexical and phonological nodes could be varied independently.⁹ Whether any such simulated deficits in lexical and phonological processing would account for the pattern of STM results in a patient remains uncertain.

⁸The data in the figures are the model's results for error types at the "lexical" level for semantic and formal errors and at the "phonological" level for neologisms. Martin et al.'s figure 7 depicts semantic errors at the lexical level because it reflects commitment to a lexical item (N. Martin, personal communication) but the use of lexical-level data is somewhat misleading, since NC's responses necessarily involve the phonological level.

⁹An example of such a model is provided by R. Martin, Breedin, and Damian (1999), who simulated the performance of one patient, AP, in phoneme discrimination, lexical decision and repetition through rapid decay of phonemic representations in a modified version of McClelland and Rumelhart's (1981) "interactive activation" model of visual word perception. R. Martin, Breedin, et al. (1999) suggested that AP's STM limitations could have resulted from his lexical deficit, but this possibility was not explored computationally.

In summary, efforts to derive properties of STM performance from properties of lexical processing constitute new ways to account for STM phenomena. These new directions pose challenges to the WM model because of their introduction of lexical factors into STM stores. They require more work to clarify the relation between aspects of lexical processing and aspects of STM.

Cases with other STM deficits

The last set of studies we shall review are ones by R. Martin and her colleagues that have been interpreted as providing evidence for a “semantic short-term memory” system. The implication of the existence of such a system would be that the WM model is incomplete.

R. Martin, Shelton, and Yaffee (1994) compared patients AB and AE, who were matched on concrete word span. AB had a somewhat higher digit span than AE. Unlike AE, AB showed greater auditory than visual span, a recency effect, a WLE, and a PSE in auditory matching span, and no difference between word span and nonword span. AB had a larger recognition span for rhyming probes and a smaller recognition span for semantic category probes than AE. Confirming a deficit in retention of semantic information, AB had greater difficulty than AE on the second through fourth sections of the Token Test, which require integration of two prenominal adjectives with one or two nouns (e.g., *Touch the large blue square and the small red circle*), and in answering questions about items that have particular attributes (e.g., *Which is quiet, a concert or a library?*). AB’s knowledge of category membership and items’ attributes was good, but his access to semantic knowledge was slowed, as measured by long reaction times in property verification, living–nonliving decisions, and in semantic relatedness (category membership) judgement (where he was also less accurate than controls).

R. Martin et al. (1994) argue that the results constitute a double dissociation that “could not be accounted for within the single capacity view of short-term [memory] and consequently provide evidence for a multiple-capacity theory” (p. 94). They reject the possibility that AB’s STM and sentence performances could have resulted from slowed semantic access, pointing to the fact that he was faster than AE on the property verification task. However, AB was considerably slower than controls in all semantic tasks, and his STM and sentence performances may have been influenced by slowed semantic access, a possibility similar to the arguments made by R. Martin, Lesch, et al. (1999) regarding MS and by N. Martin et al. (1994, 1996) regarding NC. Neither the fact that AE was slower than AB on one semantic task, nor that AB was slower than controls on baseline simple and choice RT tasks, rules out this account.

R. Martin and He (2004) described a second semantic STM case. ML had a span of 2, showed no lexical effect, reduced rhyme and category probe spans, and an auditory but not a visual PSE. Martin and He argued this pattern pointed to both phonological and semantic STM deficits. ML tended to initiate recall with the list-initial item followed by the list-final item, which Martin and He said is consistent with N. Martin and Saffran’s (1997) observations of recall order in patients with semantic deficits. ML had scored above average on the Peabody Picture Vocabulary test and the Philadelphia naming test. On the timed category and living–nonliving judgement tasks used by R. Martin et al. (1994), ML’s accuracy was within the normal range and his RTs were similar to AB’s. The same concerns about slowed semantic processing affecting STM that apply to AB apply to ML.

Hanten and Martin (2000) contrasted two paediatric head injury cases (CB, CS, both aged 10), whose performances on tests of word comprehension (the Peabody Picture Vocabulary test and a property verification test for words) and phonological discrimination were normal, and whose span performances were low. CS showed a greater than normal lexicality effect;

an auditory but not a visual PSE; an apparent interaction of effects of estimated age of acquisition and imageability of words such that performance was normal for high-imageability words and, in that set, higher for words acquired early; and reversed recency effects. CB showed a normal lexicality effect, with poorer performance on nonword lists; an auditory but not a visual PSE; an apparent interaction of effects of estimated age of acquisition and imageability of words such that performance was at the low end of normal for low-imageability words acquired early; and reversed recency effects. Hantin and Martin argued that the lower span, reversed recency, and absent visual phonological similarity effects argued for reduced phonological storage in both cases. It seems reasonable to conclude that both children had reduced PSs (within the WM framework), although the absent visual PSE could obviously be due to a recoding problem in these young children, who preserved the auditory PSE. Hantin and Martin further argued that the fact that CB's lexicality effect was smaller than CS's and that he showed no imageability effect are "evidence of a semantic retention deficit" (p. 351). Both these arguments are weak. CB's lexicality effect was in the normal range, suggesting that CS makes more use of lexical representations than normal, not that CB has a deficit in this function. The reversed imageability effect for early-acquired words is not explained by a semantic storage deficit, which would not improve performance on low-imageability words. In addition, lexical semantic processing was assessed using untimed tests, so the possibility raised earlier that slower semantic (or lexical) access could have "bled through" to STM was not addressed.

In summary, these cases further document the effects of lexical status and semantic variables on span in patients. They fall short of showing that there is a separate semantic STM store. However, the alternative to the conclusion that there is a separate semantic STM store is that problems in semantic access affect STM performances. These cases suggest that it may be possible to account for some features of STM performance in patients by considering lexical (and semantic) processing. An obvious next step would be to model the effects of delayed semantic access on span.

Summary of neuropsychological studies

The neuropsychological literature can be broadly divided in two with respect to the theoretical positions it maintains regarding the implications of neuropsychological cases for models of STM tasks. The first, somewhat older, position is that these cases support the Baddeley WM model, with some relatively minor modifications; Baddeley often refers to neuropsychological cases as providing support for the WM model (e.g., Baddeley, 2001). The second, somewhat more recent, position is that neuropsychological cases provide evidence for the role of lexical processes in producing some aspects of STM performance and for the existence of additional STM stores (specifically, a semantic STM store), and therefore pose challenges to the WM model. Our review argues that the cases that have been taken as support for the Baddeley WM model show performances that are not easily explicable within that model or in any obvious extension of that model, and that the cases that have been presented as showing that lexical processing accounts for STM phenomena in neuropsychological cases or that additional STM stores exist run into empirical and interpretive problems. The "lexical" and "semantic STM" cases do provide evidence that lexical processing deficits may underlie some aspects of patients' STM performances, and thus reinforce models of STM that incorporate multiple levels of linguistic representation. Relating the effects of these variables to models of the mechanisms that underlie STM performance is still a challenge.

(BRIEF) CONCLUSIONS

This very brief, selective, critical review has examined the evidence from normal and neuropsychological cases for the "slave systems" in the WM model. We have argued that

neither source of data provides strong support for those systems. The remaining evidence for the slave systems is said to be patterns of neural responses, especially BOLD signal measured by fMRI, to variables in STM tasks. Reviewing this literature is beyond the scope of this paper.

We conclude by putting this aspect of the WM model into a broader context. We would say that the study of the phenomena that the slave systems are said to account for is, in a basic sense, misdirected. In our view, the phenomena in immediate serial recall and related tasks that this aspect of the WM model addresses—effects of stimulus characteristics (length, phonological confusability) and concurrent events (irrelevant speech and articulation)—are secondary features of STM. The central features of STM are largely found in the descriptive name of the task with which we began and which we, and researchers contributing to the literature, have focused on the most: *immediate serial recall*. A model of STM must, we would argue, account for the effect of the temporal period over which memory is required on performance (immediacy); the ability to represent, store, and recall serial order; and the mechanisms of recall. The WM model does not deal at all with the latter two of these features, and its answer to the first is to postulate a decaying PS for which there is little or no evidence. It is perhaps uncharitable, but nonetheless not unreasonable, to maintain that the WM slave systems deal with peripheral aspects of STM, and fail to account for even them.

T. Shallice pointed out in a review of an earlier version of this paper that the WM model is an out-of-date, first-generation, information-processing model, and that the real need is to explore second-generation computational models. We entirely agree. In our view the work that has gone into our review will be justified if the review encourages neuropsychological researchers to consider their data in relation to some of this newer work. In this connection we note that these newer models, which have been connected to the WM model, include models of STM, models that attempt to account for aspects of both STM and LTM, and models that explore the nature of short-duration memory processes in normal language comprehension and production and their role in STM. Our review has pointed to some of the points of contact between some of these models and the WM model, some of the differences between them, and some of the challenges that their integration faces.

References

- Allen R, Hulme C. Speech and language processing mechanisms in verbal serial recall. *Journal of Memory and Language*. 2006; 55:64–88.
- Baddeley A. Is working memory working? The Fifteenth Bartlett lecture. *The Quarterly Journal of Experimental Psychology*. 1998; 44A:1–31.
- Baddeley AD. Short-term memory for word sequences as a function of acoustic, semantic and formal similarity. *Quarterly Journal of Experimental Psychology*. 1966a; 18:362–365. [PubMed: 5956080]
- Baddeley AD. The influence of acoustic and semantic similarity on long-term memory for word sequences. *Journal of Verbal Learning and Verbal Behavior*. 1966b; 5:217–220.
- Baddeley AD. How does acoustic similarity influence short-term memory? *The Quarterly Journal of Experimental Psychology*. 1968; 20:249–263. [PubMed: 5683764]
- Baddeley AD. The phonological loop and the irrelevant speech effect: Some comments on Neath. *Psychonomic Bulletin & Review*. 2000a; 7(3):544–549. [PubMed: 11082863]
- Baddeley AD. The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*. 2000b; 4:417–423. [PubMed: 11058819]
- Baddeley AD. Is working memory still working? *American Psychologist*. 2001; 56:851–864. [PubMed: 11785152]
- Baddeley AD, Andrade J. Reversing the word-length effect: A comment on Caplan, Rochon, and Waters. *The Quarterly Journal of Experimental Psychology*. 1994; 47A:1047–1054.

- Baddeley AD, Gathercole S, Papagno C. The phonological loop as a language learning device. *Psychological Review*. 1998; 105(1):158–173. [PubMed: 9450375]
- Baddeley, AD.; Hitch, G. Working memory. In: Bower, GH., editor. *The psychology of learning and motivation: Advances in research and theory*. Vol. 8. New York, NY: Academic Press; 1974. p. 47-89.
- Baddeley, AD.; Hitch, GJ. Recency re-examined. In: Dornic, S., editor. *Attention and performance VI*. London, UK: Academic Press; 1977. p. 647-665.
- Baddeley AD, Hitch GJ. The recency effect: Implicit learning with explicit retrieval? *Memory and Cognition*. 1993; 21:146–155. [PubMed: 8469122]
- Baddeley AD, Lewis V, Vallar G. Exploring the articulatory loop. *Quarterly Journal of Experimental Psychology*. 1984; 36A:233–252.
- Baddeley AD, Thomson N, Buchanan M. Word length and the structure of short- term memory. *Journal of Verbal Learning and Verbal Behavior*. 1975; 14:575–589.
- Baddeley AD, Wilson B. Phonological coding and short-term memory in patients without speech. *Journal of Memory and Language*. 1985; 24:490–502.
- Basso A, Spinnler H, Vallar G, Zanobio ME. Left hemisphere damage and selective impairment of auditory verbal short-term memory: A case study. *Neuropsychologia*. 1982; 20:263–274. [PubMed: 7121794]
- Beaman, CP.; Neath, I.; Surprenant, AM. Phonological similarity effects without a phonological store: An individual differences model. In: McNamara, DS.; Trafton, G., editors. *Proceedings of the Twenty Ninth Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Erlbaum; 2007. p. 89-94.
- Besner D. Phonology, lexical access in reading, and articulatory suppression: A critical review. *The Quarterly Journal of Experimental Psychology*. 1987; 39A:467–478.
- Bireta TJ, Neath I, Surprenant AM. The syllable-based word length effect and stimulus set specificity. *Psychonomic Bulletin & Review*. 2006; 13:434–438. [PubMed: 17048727]
- Bjork RA, Whitten WB. Recency-sensitive retrieval processes in long term free recall. *Cognitive Psychology*. 1974; 6:173–189.
- Bourassa DC, Besner D. Beyond the articulatory loop: A semantic contribution to serial order recall of subspan lists. *Psychonomic Bulletin and Review*. 1994; 1:122–125. [PubMed: 24203421]
- Bregman, AS. *Auditory scene analysis: The perceptual organisation of sound*. Cambridge, MA: MIT Press; 1990.
- Brown GDA, Neath I, Chater N. A ratio model of scale-invariant memory and identification. *Psychological Review*. 2007; 114:539–576. [PubMed: 17638496]
- Brown J. Some tests of the decay theory of immediate memory. *Quarterly Journal of Experimental Psychology*. 1958; 10:12–21.
- Burgess N, Hitch GJ. Towards a network model of the articulatory loop. *Journal of Memory and Language*. 1992; 31:429–460.
- Burgess N, Hitch GJ. Memory for serial order: A network model of the phonological loop and its timing. *Psychological Review*. 1999; 106:551–581.
- Caplan D, Rochon E, Waters GS. Articulatory and phonological determinants of word length effects in span tasks. *Quarterly Journal of Experimental Psychology*. 1992; 45A:177–192. [PubMed: 1410554]
- Caplan D, Vanier M, Baker C. A case study of reproduction conduction aphasia I: Word production. *Cognitive Neuropsychology*. 1986; 3(1):99–128.
- Caplan, D.; Waters, GS. Short-term memory and language comprehension: A critical review of the neuropsychological literature. In: Vallar, G.; Shallice, T., editors. *Neuropsychological impairments of short-term memory*. Cambridge, UK: Cambridge University Press; 1990. p. 337-389.
- Caplan D, Waters G. Articulatory length and phonological similarity in span tasks: A reply to Baddeley and Andrade. *Quarterly Journal of Experimental Psychology*. 1994; 47A:1055–1062. [PubMed: 7809398]

- Caramazza A. On drawing inferences about the structure of normal cognitive systems from the analysis of patterns of impaired performance: The case for single-patient studies. *Brain & Cognition*. 1986; 5:41–66. [PubMed: 3954906]
- Colle HA. Auditory encoding in visual short-term recall: Effects of noise intensity and spatial location. *Journal of Verbal Learning and Verbal Behavior*. 1980; 19:722–735.
- Colle HA, Welsh A. Acoustic masking in primary memory. *Journal of Verbal Learning and Verbal Behavior*. 1976; 15:17–32.
- Conrad R. Acoustic confusions and memory span for words. *Nature*. 1963; 197:1029–1030. [PubMed: 14022639]
- Conrad R. Acoustic confusions in immediate memory. *British Journal of Psychology*. 1964; 55:75–84.
- Conrad R, Hull AJ. Information, acoustic confusion and memory span. *British Journal of Psychology*. 1964; 55:429–432. [PubMed: 14237884]
- Cowan N. Activation, attention and short-term memory. *Memory and Cognition*. 1993; 21:162–167. [PubMed: 8469124]
- Cowan, N. Attention and memory: An integrated framework. New York, NY: Oxford University Press; 1995. Oxford Psychology Series (No. 26)
- Cowan N, Baddeley AD, Elliott EM, Norris J. List composition and the word length effect in immediate recall: A comparison of localist and globalist assumptions. *Psychonomic Bulletin & Review*. 2003; 10:74–79. [PubMed: 12747492]
- Cowan N, Day L, Saults JS, Keller TA, Johnson T, Flores L. The role of verbal output time in the effects of word length on immediate memory. *Journal of Memory and Language*. 1992; 31:1–17.
- Cowan N, Elliott EM, Saults JG, Nugent LD, Bomb P, Hismjatullina A. Rethinking speed theories of cognitive development: Increasing the rate of recall without affecting accuracy. *Psychological Science*. 2006; 17:67–73. [PubMed: 16371146]
- Cowan, N.; Kail, R. Covert processes and their development in short-term memory. In: Gathercole, SE., editor. *Models of short-term memory*. Hove, UK: Psychology Press; 1996.
- Cowan N, Nugent LD, Elliott EM, Geer T. Is there a temporal basis of the word length effect? A response to Service (1998). *Quarterly Journal of Experimental Psychology*. 2000; 53A:647–660. [PubMed: 10994222]
- Craik FIM. Two components in free recall. *Journal of Verbal Learning and Verbal Behavior*. 1968a; 7:996–1004.
- Craik FIM. Types of error in free recall. *Psychonomic Science*. 1968b; 10:353–354.
- Craik FIM. The fate of primary memory items in free recall. *Journal of Verbal Learning and Verbal Behavior*. 1970; 9:143–148.
- Craik FIM, Levy BA. Semantic and acoustic information in primary memory. *Journal of Experimental Psychology*. 1970; 86:77–82.
- Crossman, ERFW. *Information theory*. London, UK: Butterworth; 1959. Information and order in immediate memory.
- Damasio H, Damasio AR. The anatomical basis of conduction aphasia. *Brain*. 1980; 103:337–350. [PubMed: 7397481]
- Dell GS. A spreading activation theory of retrieval in sentence production. *Psychological Review*. 1986; 93:283–321. [PubMed: 3749399]
- Dell GS, Martin N, Schwartz MF. A case-series test of the interactive two-step model of lexical access: Predicting word repetition from picture naming. *Journal of Memory and Language*. 2007; 56:490–520. [PubMed: 21085621]
- Dell GS, O'Seaghdha PG. Mediated and convergent lexical priming in language production: A comment on Levelt et al. (1990). *Psychological Review*. 1991; 98:604–614. [PubMed: 1961775]
- Dell GS, Schwartz MF, Martin N, Saffran EM, Gagnon DA. Lexical access in aphasic and non-aphasic speakers. *Psychological Review*. 1997; 104(4):801–838. [PubMed: 9337631]
- Della Sala S, Logie RH. Impairments of methodology and theory in cognitive neuropsychology: A case for rehabilitation? *Neuropsychological Rehabilitation*. 1997; 7:367–385.

- Della Sala S, Logie RH, Trivelli C, Cubelli R, Marchetti C. Dissociation between recency and span: Neuropsychological and experimental evidence. *Neuropsychology*. 1998; 12(4):533–545. [PubMed: 9805323]
- Farley LA, Neath I, Allbritton DW, Surprenant AM. Irrelevant speech effects and sequence learning. *Memory & Cognition*. 2007; 35:156–165. [PubMed: 17533889]
- Glanzer M, Cunitz AR. Two storage mechanisms in free recall. *Journal of Verbal Learning and Verbal Behavior*. 1966; 5:351–360.
- Glanzer M, Koppenaal L, Nelson R. Effects of relations between words on short- term storage and long-term storage. *Journal of Verbal Learning and Verbal Behavior*. 1972; 11:403–416.
- Glanzer M, Razel M. The size of the unit in short-term storage. *Journal of Verbal Learning and Verbal Behavior*. 1974; 13(1):114–131.
- Glenberg AM, Bradley M, Stecenson J, Kraus T, Tzachuk M, Gretz A, et al. A two process account of long term serial position effects. *Journal of Experimental Psychology: Learning, Memory and Cognition*. 1980; 6:335–369.
- Goodglass H, Winfield A, Hyde MR, Gleason JB, Bowles NL, Gallagher RE. The importance of word-initial phonology: Error patterns in prolonged naming efforts by aphasic patients. *Journal of the International Neuropsychological Society*. 1997; 3:128–138. [PubMed: 9126854]
- Greene RL. Sources of recency effects in free recall. *Psychological Bulletin*. 1986; 99:221–228.
- Guérard K, Jalbert A, Neath I, Surprenant AM, Bireta TJ. Irrelevant tapping and the acoustic confusion effect: The role of spatial complexity. *Experimental Psychology*. 2009; 56:367–374. [PubMed: 19447753]
- Hall LW, Wilson KP, Humphreys MS, Tinzmann MB, Bowyer PM. Phonemic-similarity effects in good versus poor readers. *Memory and Cognition*. 1983; 11:520–527. [PubMed: 6656611]
- Hanley JR, Broadbent C. The effect of unattended speech on serial recall following auditory presentation. *British Journal of Psychology*. 1987; 78:287–297.
- Hanten G, Martin R. Contributions of phonological and semantic short-term memory to sentence processing: Evidence from two cases of closed head injury in children. *Journal of Memory and Language*. 2000; 43:335–361.
- Henson RNA. Short-term memory for serial order. The start-end model. *Cognitive Psychology*. 1998; 36:73–137. [PubMed: 9721198]
- Henson RNA, Norris DG, Page MPA, Baddeley AD. Unchained memory: Error patterns rule out chaining models of immediate serial recall. *The Quarterly Journal of Experimental Psychology*. 1996; 49A:80–115.
- Hinton, GE.; Plaut, DC. *Proceedings of the 9th Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Lawrence Erlbaum Associates; 1987. Using fast weights to deblur old memories; p. 177-186.
- Howard, D.; Franklin, S. *Missing the meaning? A cognitive neuropsychological analysis of single word processing in an aphasic patient*. Cambridge, MA: MIT Press; 1988.
- Howard, D.; Franklin, S. *Memory without rehearsal*. In: Vallar, G.; Shallice, T., editors. *Neuropsychological impairments of short-term memory*. Cambridge, UK: Cambridge University Press; 1990.
- Howard D, Nickels LA. Separating input and output phonology: Semantic, phonological and orthographic effects in short-term memory impairment. *Cognitive Neuropsychology*. 2005; 22:42–77. [PubMed: 21038240]
- Hulme C, Maughan S, Brown G. Memory for familiar and unfamiliar words: Evidence for a long-term memory contribution to short-term span. *Journal of Memory and Language*. 1991; 30:685–701.
- Hulme C, Newton P, Cowan N, Stuart G, Brown G. Think before you speak: Pause, memory search and trace reintegration processes in verbal memory span. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 1999; 25:447–463.
- Hulme C, Surprenant AM, Bireta TJ, Stuart G, Neath I. Abolishing the word-length effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 2004; 30:98–106.
- Jalbert A, Neath I, Bireta TJ, Surprenant AM. When does length cause the word length effect? *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 2011; 37:338–353.

- Jalbert A, Neath I, Surprenant AM. Does length or neighbourhood size cause the word length effect? *Memory & Cognition*. 2011; 39:1198–1210. [PubMed: 21461875]
- Johnston RS. Phonological coding in dyslexic readers. *British Journal of Psychology*. 1982; 73:455–460. [PubMed: 7171920]
- Johnston RS, Rugg ME, Scott T. Phonological similarity effects, memory span and developmental reading disorders: The nature of the relationship. *British Journal of Psychology*. 1987; 78:205–211. [PubMed: 3594090]
- Jones D, Hughes RW, Macken WJ. Perceptual organisation masquerading as phonological storage: Further support for a perceptual-gestural view of short-term memory. *Journal of Memory and Language*. 2006; 54:265–281.
- Jones, DM. Objects, streams and threads of auditory attention. In: Baddeley, AD.; Weiskrantz, L., editors. *Attention: Selection, awareness and control*. Oxford, UK: Oxford University Press, Clarendon Press; 1993. p. 87-104.
- Jones DM, Alford D, Bridges A, Tremblay S, Macken WJ. Organisational factors in selective attention: The interplay of acoustic distinctiveness and auditory streaming in the irrelevant sound effect. *Journal of Experimental Psychology: Learning, Memory and Cognition*. 1999; 25:464–473.
- Jones DM, Hughes RW, Macken WJ. The phonological store abandoned. *Quarterly Journal of Experimental Psychology*. 2007; 60:505–511.
- Jones DM, Hughes RW, Macken WJ. Auditory distraction and serial memory: The avoidable and the ineluctable. *Noise & Health*. 2010; 12:201–209. [PubMed: 20871174]
- Jones DM, Macken WJ. Irrelevant tones produce an irrelevant speech effect: Implications for phonological coding in working memory. *Journal of Experimental Psychology: Learning, Memory and Cognition*. 1993; 19:369–381.
- Jones DM, Macken WJ. Organisational factors in the effect of irrelevant speech: The role of spatial location and timing. *Memory and Cognition*. 1995a; 23:192–200. [PubMed: 7731364]
- Jones DM, Macken WJ. Phonological similarity in the irrelevant speech effect. Within- or between-stream similarity? *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 1995b; 21:103–115.
- Jones DM, Macken WJ, Murray AC. Disruption of visual short-term memory by changing-state auditory stimuli: The role of segmentation. *Memory and Cognition*. 1993; 21:318–328. [PubMed: 8316094]
- Kohn SE, Wingfield A, Menn L, Goodglass H, Gleason JB, Hyde MR. Lexical retrieval: The tip-of-the-tongue phenomenon. *Applied Psycholinguistics*. 1987; 8:245–266.
- Knott RA, Patterson KE, Hodges JR. Lexical and semantic binding effects in short-term memory: Evidence from semantic dementia. *Cognitive Neuropsychology*. 1997; 14(8):1165–1216.
- Larsen J, Baddeley AD. Disruption of verbal STM by irrelevant speech, articulatory suppression and manual tapping: Do they have a common source? *Quarterly Journal of Experimental Psychology*. 2003; 56A:1249–1268. [PubMed: 14578082]
- Levelt, WJM. *Speaking: From intention to articulation*. Cambridge, MA: MIT Press; 1989.
- Lewandowsky S, Oberauer K. The word-length effect provides no evidence for decay in short-term memory. *Psychonomic Bulletin and Review*. 2008; 15(5):875–888. [PubMed: 18926980]
- Logie RH, Sala SD, Laiacona M, Chalmers P, Wynn V. Group aggregates and individual reliability: The case of verbal short-term memory. *Memory & Cognition*. 1996; 24:305–321. [PubMed: 8718765]
- Lovatt P, Avons SE, Masterson J. The word-length effect and disyllabic words. *The Quarterly Journal of Experimental Psychology*. 2000; 53A(1):1–22. [PubMed: 10718061]
- Lovatt P, Avons SE, Masterson J. Output decay in immediate serial recall: Speech time revisited. *Journal of Memory and Language*. 2002; 46:227–243.
- Luce, RD. *Individual choice behavior: A theoretical analysis*. New York, NY: Wiley; 1959.
- Luce, RD. Detection and recognition. In: Luce, RD.; Bush, RR.; Galanter, E., editors. *Handbook of mathematical psychology*. New York, NY: Wiley; 1963. p. 103-189.

- Macken WJ, Jones DM. Functional characteristics of the inner voice and the inner ear: Single or double agency? *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 1995; 21:436–448.
- Macken WJ, Jones DM. Reification of phonological storage. *Quarterly Journal of Experimental Psychology*. 2003; 56A:1279–1288. [PubMed: 14578084]
- Macken WJ, Phelps F, Jones DM. What causes auditory distraction? *Psychonomic Bulletin and Review*. 2009; 16:139–145. [PubMed: 19145024]
- Mann VA, Liberman IY, Shankweiler D. Children's memory for sentences and word strings in relation to reading ability. *Memory & Cognition*. 1980; 8:329–335. [PubMed: 7421573]
- Martin N, Ayala J. Measurements of auditory-verbal STM in aphasia: Effects of task, item and word processing impairment. *Brain and Language*. 2004; 89:464–483. [PubMed: 15120538]
- Martin N, Gupta P. Exploring the relationship between word processing and verbal STM: Evidence from associations and dissociations. *Cognitive Neuropsychology*. 2004; 21:213–228. [PubMed: 21038201]
- Martin N, Saffran EM. Language and auditory-verbal short-term memory impairments: Evidence for common underlying processes. *Cognitive Neuropsychology*. 1997; 14:641–682.
- Martin N, Saffran EM. Language and auditory-verbal short-term memory impairments: Evidence for common underlying processes. *Cognitive Neuropsychology*. 1997; 14(5):641–682.
- Martin N, Saffran EM, Dell GS. Recovery in deep dysphasia: Evidence for a relation between auditory-verbal STM capacity and lexical errors in repetition. *Brain and Language*. 1996; 52:83–113. [PubMed: 8741977]
- Martin N, Saffran EM, Dell GS, Schwartz M. Origins of paraphasias in deep dysphasia: Testing the consequences of a decay impairment to an interactive spreading activation model of lexical retrieval. *Brain and Language*. 1994; 47:609–660. [PubMed: 7859057]
- Martin RC, Breedin S, Damian M. The relation of phoneme processing, lexical access and short-term memory: A case study and interactive activation account. *Brain and Language*. 1999; 70:437–482. [PubMed: 10600229]
- Martin RC, He T. Semantic short-term memory and its role in sentence processing: A replication. *Brain and Language*. 2004; 89:76–82. [PubMed: 15010239]
- Martin RC, Lesch M, Bartha M. Independence of input and output phonology in word processing and short-term memory. *Journal of Memory and Language*. 1999; 41:3–29.
- Martin RC, Shelton J, Yaffee L. Language processing and working memory: Neuropsychological evidence for separate phonological and semantic capacities. *Journal of Memory and Language*. 1994; 33:83–111.
- McClelland JL, Rumelhart DE. An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review*. 1981; 88:375–407.
- Miceli, G.; Silveri, MC.; Caramazza, A. The role of the phoneme-to-grapheme conversion system and of the graphemic output buffer in writing: Evidence from an Italian case of pure dysgraphia. In: Coltheart, M.; Sartori, G.; Job, R., editors. *Cognitive neuropsychology of language*. Hillsdale, NJ: Lawrence Erlbaum Associates; 1986. p. 235–252.
- Miller GA. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*. 1956; 63(2):81–97. [PubMed: 13310704]
- Mueller ST, Seymour TL, Kieras DE, Meyer DE. Theoretical implications of articulatory duration, phonological similarity, and phonological complexity in verbal working memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*. 2003; 29:1353–1380.
- Murdock BB Jr. TODAM2: A model for the storage and retrieval of item, associative, and serial-order information. *Psychological Review*. 1993; 100:183–203. [PubMed: 8483981]
- Murdock BB Jr. Developing TODAM: Three models for serial order information. *Memory and Cognition*. 1995; 23:631–645. [PubMed: 7476248]
- Neath I, Bireta TJ, Surprenant AM. The time-based word length effect and stimulus set specificity. *Psychonomic Bulletin and Review*. 2003; 10:430–434. [PubMed: 12921420]
- Neath I, Guérard K, Jalbert A, Bireta TJ, Surprenant AM. Irrelevant speech effects and statistical learning. *Quarterly Journal of Experimental Psychology*. 2009; 62:1551–1559.

- Nickels L, Howard D. Phonological errors in aphasic naming: Comprehension, monitoring and lexicality. *Cortex*. 1995; 31(2):209–237. [PubMed: 7555004]
- O'Connor N, Hermelin B. Seeing and hearing and space and time. *Perception and Psychophysics*. 1972; 11:46–48.
- Page, M.; Henson, RN. Models of short-term memory: Modelling immediate serial recall of verbal material. In: Andrade, J., editor. *Working memory in perspective*. Hove, UK: Psychology Press; 2001. p. 177-198.
- Page MPA, Norris D. The primacy model: A new model of immediate serial recall. *Psychological Review*. 1998; 105:761–781. [PubMed: 9830378]
- Page MPA, Norris DG. The irrelevant sound effect: What needs modeling and a tentative model. *Quarterly Journal of Experimental Psychology*. 2003; 56A:1289–1300. [PubMed: 14578085]
- Peterson LR, Peterson MJ. Short-term retention of individual verbal items. *Journal of Experimental Psychology*. 1959; 58:193–198. [PubMed: 14432252]
- Richardson JTE, Baddeley AD. The effect of articulatory suppression in free recall. *Journal of Verbal Learning and Verbal Behavior*. 1975; 14:623–629.
- Ricker T, AuBuchon AM, Cowan N. Working memory. *Wiley Interdisciplinary Reviews: Cognitive Science*. 2010; 1:573–585.
- Rochon E, Caplan D, Waters GS. Short-term memory processes in patients with apraxia of speech: Implications for the nature and structure of the auditory verbal short-term memory system. *Journal of Neurolinguistics*. 1991; 5:231–264.
- Roodenrys S, Hulme C, Lethbridge A, Hinton M, Nimmo LM. Word-frequency and phonological-neighbourhood effects on verbal short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 2002; 28:1019–1034.
- Saffran, EM.; Martin, N. Neuropsychological evidence for lexical involvement in short-term memory. In: Vallar, G.; Shallice, T., editors. *Neuropsychological impairments of short-term memory*. Cambridge, UK: Cambridge University Press; 1990. p. 145-166.
- Saffran JR, Aslin RN, Newport EL. Statistical learning by 8-month-old infants. *Science*. 1996; 274:1926–1928. [PubMed: 8943209]
- Saito S. Influence of articulatory suppression and memory updating on phonological similarity effect. *Japanese Journal of Psychology*. 1993; 64:289–295. In Japanese with English abstract. [PubMed: 8301894]
- Saito S. What effect can rhythmic finger tapping have on the phonological similarity effect? *Memory and Cognition*. 1994; 22(2):181–187. [PubMed: 8035694]
- Salamé P, Baddeley AD. Disruption of short-term memory by unattended speech: Implications for the structure of working memory. *Journal of Verbal Learning and Verbal Behavior*. 1982; 21:150–164.
- Salamé P, Baddeley AD. Phonological factors in STM: Similarity and the unattended speech effect. *Bulletin of the Psychonomic Society*. 1986; 24:263–265.
- Service E. The effect of word length on immediate serial recall depends on phonological complexity, not articulatory duration. *Quarterly Journal of Experimental Psychology*. 1998; 51A:283–304.
- Shah P, Miyake A. The separability of working memory resources for spatial thinking and language processing: An individual differences approach. *Journal of Experimental Psychology: General*. 1996; 125:4–27. [PubMed: 8851737]
- Shallice, T. On the contents of primary memory. In: Rabbit, PMA.; Domic, S., editors. *Attention and performance*. Vol. 5. London: Academic Press; 1975. p. 269-280.
- Shallice, T. *From neuropsychology to mental structure*. Cambridge, UK: Cambridge University Press; 1988.
- Shallice T, Butterworth B. Short-term memory impairment and spontaneous speech. *Neuropsychologia*. 1977; 15:729–735. [PubMed: 600367]
- Shallice, T.; Vallar, G. The impairment of auditory-verbal short-term storage. In: Vallar, G.; Shallice, T., editors. *Neuropsychological impairments of short-term memory*. Cambridge, UK: Cambridge University Press; 1990. p. 11-53.

- Shallice T, Warrington EK. Independent functioning of verbal memory stores: A neuropsychological study. *Quarterly Journal of Experimental Psychology*. 1970; 22:261–273. [PubMed: 5431401]
- Shallice T, Warrington EK. Auditory-verbal short-term memory impairment and conduction aphasia. *Brain and Language*. 1977; 4(4):479–491. [PubMed: 922463]
- Shankweiler D, Liberman IY, Mark LS, Fowler CA, Fischer FW. The speech code and learning to read. *Journal of Experimental Psychology: Human Learning & Memory*. 1979; 5:531–545.
- Sperling G. Successive approximations to a model for short-term memory. *Acta Psychologica*. 1967; 27:285–292. [PubMed: 6062221]
- Sperling, G.; Spelman, RG. Acoustic similarity and auditory short term memory: Experiments and a model. In: Norman, DA., editor. *Models of human memory*. New York, NY and London, UK: Academic Press; 1970. p. 151
- Surprenant AM, Neath I, Bireta TJ, Allbritton DW. Directly assessing the relationship between irrelevant speech and irrelevant tapping. *Canadian Journal of Experimental Psychology*. 2008; 62:141–149. [PubMed: 18778142]
- Thorn ASC, Frankish CR. Long-term knowledge effects on serial recall of nonwords are not exclusively lexical. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 2005; 31:729–735.
- Tolan GA, Tehan G. Is spoken duration a sufficient explanation of the word length effect? *Memory*. 2005; 13:372–379. [PubMed: 15948623]
- Tree JJ, Longmore C, Besner D. Orthography, phonology, short-term memory and the effects of concurrent articulation on rhyme and homophony judgements. *Acta Psychologica*. 2011; 136:11–19. [PubMed: 20965480]
- Turk-Browne NB, Jungé J, Scholl BJ. The automaticity of visual statistical learning. *Journal of Experimental Psychology: General*. 2005; 134:552–564. [PubMed: 16316291]
- Tzeng OJL. Positive recency effect in delayed free recall. *Journal of Verbal Learning and Verbal Behavior*. 1973; 12:436–439.
- Vallar G, Baddeley AD. Phonological short-term store and sentence processing. *Cognitive Neuropsychology*. 1987; 4:417–438.
- Vallar G, Cappa SF. Articulation and verbal short-term memory. Evidence from anarthria. *Cognitive Neuropsychology*. 1987; 4:55–78.
- Vallar G, Di Betta AM, Silveri MC. The phonological short-term store-rehearsal system: Patterns of impairment and neural correlates. *Neuropsychologia*. 1997; 35:795–812. [PubMed: 9204486]
- Vallar G, Papagno C. Phonological short-term store and the nature of the recency effect. Evidence from neuropsychology. *Brain and Cognition*. 1986; 5:428–442. [PubMed: 3580186]
- Vallar, G.; Papagno, C. Neuropsychological impairments of short-term memory. In: Baddeley, AD.; Wilson, BA.; Watts, F., editors. *Handbook of memory disorders*. Chichester, UK: John Wiley; 1995. p. 135–165.
- Vallar, G.; Papagno, C. Neuropsychological impairments of verbal short-term memory. In: Baddeley, A.; Wilson, B.; Kopelman, M., editors. *Handbook of memory disorders*. 2. Chichester, UK: John Wiley; 2002. p. 249–270.
- Walker I, Hulme C. Concrete words are easier to recall than abstract words: Evidence for a semantic contribution to short-term serial recall. *Journal of Experimental Psychology: Learning, Memory, & Cognition*. 1999; 25:1256–1271.
- Ward, G. A critique of the working memory model. In: Andrade, J., editor. *Working memory in perspective*. Hove, UK: Psychology Press; 2001. p. 219–239.
- Warrington EK, Logue V, Pratt RTC. The anatomical localisation of selective impairment of auditory verbal short-term memory. *Neuropsychologia*. 1971; 9:377–387. [PubMed: 5164373]
- Waters GS, Rochon E, Caplan D. The role of high-level speech planning in rehearsal: Evidence from patients with apraxia of speech. *Journal of Memory and Language*. 1992; 31:54–73.
- Watkins MJ. Locus of the modality effect in free recall. *Journal of Verbal Learning and Verbal Behavior*. 1972; 11:644–648.
- Watkins MJ, Watkins OC, Crowder RG. The modality effect in free and serial recall as a function of phonological similarity. *Journal of Verbal Learning and Verbal Behavior*. 1974; 13:430–447.

- Watkins MJ, Watkins OJ. The postcategorical status of the modality effect in serial recall. *Journal of Experimental Psychology*. 1973; 99:226–230.
- Wickelgren WA. Short-term memory for phonemically similar lists. *American Journal of Psychology*. 1965; 78:567–574. [PubMed: 5839924]
- Wickelgren WA. Associative intrusions in short-term recall. *Journal of Experimental Psychology*. 1966; 72:853–858.
- Zhang W, Feng L. The visual recognition and capacity of STM for Chinese disyllabic words. *Acta Psychologica Sinica*. 1990; 22:383–390.

Figure 1.

Serial position curves for PV and controls. Top row of figures: free recall; Second row of figures: Recall from end. Left panels: auditory presentation; Right panels: visual presentation; Dotted lines: controls; Straight lines: PV. Reprinted from *Brain and Cognition*, 5(4), Giuseppe Vallar & Costanza Papagno, Phonological short-term store and the nature of the recency effect: Evidence from neuropsychology, pages 428–442, Copyright 1986, with permission from Elsevier.

Figure 2.

Performance and simulation of NC's recall of low- and high-imageable words in first and second position in two word lists and after decay. Left panels: NC's performance on high-imageability (left) and low-imageability (right) words in first (top panel) and second (middle panel) position of two word lists, and after a filled delay (bottom panel). Right panels: Simulation. Top panel: Simulated performance on first word in two word lists with high-imageability (left) and low-imageability (right) words. Middle panel: Simulated performance on second word in two word lists with high-imageability (left) and low-imageability (right) words; Bottom panel: Martin et al.'s (1996) figure 7, showing effect of imageability on semantic errors collapsed over time. S = semantic error; F = formal paraphasia; N = neologism. Reprinted from *Brain and Language*, 52(1), Nadine Martin, Eleanor M. Saffran, & Gary S. Dell, Recovery in deep dysphasia: Evidence for a relation between auditory-verbal STM capacity and lexical errors in repetition, pages 83–113, Copyright 1996, with permission from Elsevier.