BEHAVIORAL AND BRAIN SCIENCES (1999) **22,** 77–126
Printed in the United States of America

Verbal working memory and sentence comprehension

David Caplan
Neuropsychology Laboratory, Department of Neurology, Massachusetts General Hospital, Boston, MA 02114
caplan@helix.mgh.harvard.edu

Gloria S. Waters
Department of Communication Disorders, Boston University, Boston, MA 02215
gwaters@bu.edu

Abstract: This target article discusses the verbal working memory system used in sentence comprehension. We review the concept of working memory as a short-duration system in which small amounts of information are simultaneously stored and manipulated in the service of accomplishing a task. We summarize the argument that syntactic processing in sentence comprehension requires such a storage and computational system. We then ask whether the working memory system used in syntactic processing is the same as that used in verbally mediated tasks that involve conscious controlled processing. Evidence is brought to bear from various sources: the relationship between individual differences in working memory and individual differences in the efficiency of syntactic processing; the effect of concurrent verbal memory load on syntactic processing; and syntactic processing in patients with poor short-term memory, patients with poor working memory, and patients with aphasia. Experimental results from these normal subjects and patients with various brain lesions converge on the conclusion that there is a specialization in the verbal working memory system for assigning the syntactic structure of a sentence and using that structure in determining sentence meaning that is separate from the working memory system underlying the use of sentence meaning to accomplish other functions. We present a theory of the divisions of the verbal working memory system and suggestions regarding its neural basis.

Keywords: memory; sentence comprehension; verbal working memory; working memory

Introduction

Research in psychology has provided considerable evidence for a division between long-term memory, in which memories of large numbers of facts and autobiographical events are maintained for up to years, and short-term memory, which is capable of retaining small amounts of information for very short periods of time (Squire & Zola-Morgan 1991). Baddeley and his colleagues introduced the powerful concept that the appropriate way to characterize short-term memory is as a "working memory" system (Baddeley 1976; 1986; 1995; Baddeley & Hitch 1974). Working memory is conceived of as a short-duration, limited-capacity memory system capable of simultaneously storing and manipulating information in the service of accomp-lishing a task (Baddeley 1995). Appeal to the notion of a limited-capacity working memory system (or to equivalent concepts, such as "processing resources") to account for features of human cognitive performance is widespread in cognitive psychology, with respect to both normal functions (Just & Carpenter 1992) and the abilities of subjects with developmental and acquired cognitive disorders (Gathercole & Baddeley 1993).

Baddeley and his colleagues proposed the first model of the functional architecture of human working memory. In their model, working memory is made up of three main components – the central executive, the articulatory loop, and the visual-spatial scratch pad (Baddeley 1986; 1990a; Baddeley & Hitch 1994; Baddeley et al. 1986; Gathercole & Baddeley 1993). The articulatory loop and the visual-spatial scratch pad are "slave systems" in which verbal and visual information, respectively, is stored when the central executive is overloaded. One can conceive of these components as responsible for maintaining short-term information availability. The central executive is the workhorse and mastermind of human cognition. It allocates attention to a task and performs information storage and computational functions within a given task.

To determine the working memory requirements of a task, Baddeley's theory of the functional architecture of

David Caplan, M.D., Ph.D., is Professor of Neurology at the Harvard Medical School. He is the author of over 100 scientific publications in the area of language and language disorders, including *Language: Structure, Processing and Disorders* (MIT Press, 1992) and *Neurolinguistics and Linguistic Aphasiology* (Cambridge University Press, 1987).

Gloria Waters is a Professor in the Department of Communication Disorders at Boston University. She is the author of over 50 publications in the area of language and memory. Her current work is on the effect of normal aging and various neuropathological conditions on language comprehension.

 0140-525X/99 $12.50

working memory must be supplemented by specific models of the computational demands that individual cognitive processes make on the central executive. Current models of cognitive processes provide such measures. Implemented models provide specific quantitative measurements of the computational and storage demands of a task. For instance, in procedural models (e.g., Lewis 1996a), a measure of the working memory requirements of a computation might be the number of procedures required to reach a subgoal and the number of elements maintained in each procedure [see also Newell: Précis of "Unified Theories of Cognition," *BBS* 15(3) 1992]. In contrast, measures of working memory requirements derived from neural net models would be quite different and might consist in the number of steps needed to gravitate toward an attractor in a neural network (Tabor et al. 1997). The working memory demands exerted at a given point of processing in a task are usually taken to be the sum of the working memory requirements of the functions active at that point in the task (Just & Carpenter 1992). Many valuable models of cognitive phenomena are not implemented, but they still provide guides to the relative complexity of one operation, or set of operations, compared to another (Gibson & Thomas 1996). Research into the role of working memory in cognitive processes regularly appeals to all these types of models to provide the basis for determination of the processing demands of a task and of those demands at a particular point of processing within a task.

Many basic questions about working memory in human cognition remain unanswered. This target article examines possible specializations in working memory. There is considerable evidence for a division of the central executive of the working memory system into visual and verbal components (Shah & Miyake 1996). We focus on a finer division, the possibility that the verbal working memory system is composed of subsystems devoted to different types of verbal tasks.

Our particular interest is in the distinction between the extraction of meaning from a linguistic signal, which we call "interpretive processing," and in the use of that meaning to accomplish other tasks, such as storing information in long-term semantic memory, reasoning, planning actions, and other functions, which we call "post-interpretive processing." By "interpretive processing," we refer to the processes of recognizing words and appreciating their meanings and syntactic features; constructing syntactic and prosodic representations; and assigning thematic roles, focus, and other aspects of propositional and discourse-level semantics (see below for further discussion). Many linguists and psycholinguists have argued that the processes involved in interpretive processing are distinct from those involved in other verbally mediated functions (e.g., Fodor 1983; Forster 1979; Frazier 1990). Arguments regarding the "modularity" of interpretive processing have centered largely on the issue of what types of information are used in the initial determination of linguistic form and meaning (MacDonald 1997). We will address the question of a cognitive specialization for interpretive processing from the point of view of the structure of working memory.

Our empirical focus is on the process of sentence comprehension. We will review the results of a variety of experiments suggesting that the working memory system that is called on in interpretive processing at the sentence level – assigning the syntactic structure of a sentence and using that structure to determine the meaning of the sentence – constitutes a separate subsystem within verbal working memory. After presenting this evidence, we will set the topic in a broader context and advance the hypothesis that this subsystem of verbal working memory is involved in the set of related operations responsible for identifying linguistic elements and structures and determining the preferred literal meaning of an utterance.

This paper is divided into four main sections. First, we will review the evidence that processing resources are involved in sentence comprehension. We will then review evidence from normal subjects that bears on the nature of the working memory system involved in syntactic processing in sentence comprehension, followed by a presentation of similar evidence from several groups of patients with central nervous system (CNS) disease. We will end this target article with a statement of our theory, including its neurological aspects.

1. Sentence comprehension and working memory

There are several pieces of evidence indicating that comprehension of sentences requires verbal working memory resources. To begin with anecdotal evidence, consider the following sentence:

1. The man that the woman that the child hugged kissed laughed.

In sentence (1), most readers have trouble assigning thematic roles (who did what to whom). They can assign thematic roles much more easily in the two sentences that combine to form it, shown in (2):

2a. The man that the woman kissed laughed.
b. The woman that the child hugged kissed the man.

The trouble that normal English users have understanding sentence (1) is thought to arise because they do not have sufficient working memory to retain the intermediate products of computation that are produced in building the complex syntactic structure of this sentence (Chomsky 1957; Gibson 1997; Johnson 1996).

A variety of implemented models of syntactic processing (parsing) have been developed. These models produce measurements of the computational demands of assigning syntactic structure and using it to determine aspects of sentence meaning. These models have taken many forms, ranging from deterministic parsers associated with Chomsky's (1981; 1986; 1993) model of syntactic representations (Berwick & Weinberg 1984) through Abney's (1989) principles-and-parameters parser, Johnson's (1996) logical parser, connectionist models (Tabor et al. 1997), and various hybrid models such as Just and Carpenter's CC-READER (Just & Carpenter 1992). Nonimplemented models of how sentences are parsed have also been developed (e.g., Frazier & Clifton 1996; Gibson, in press). Although there are considerable differences among the types of operations and complexity metrics found in these models, there is a remarkable degree of similarity in the measurements they produce of the relative complexity of the parsing process in many different types of sentences.

With partial exploitation of these models, experimental results have provided evidence beyond intuitions that syntactic processing in sentence comprehension requires the allocation of processing resources. Considerable research has found evidence that sentences that have more complex syntactic structures are more difficult and time consuming to understand (see papers in MacDonald 1997). The evidence that syntactic structural complexity is associated with increased difficulty in sentence processing extends to measurements made internal to the sentence-processing process. Eye-fixation durations, self-paced word-by-word and phrase-by-phrase reading and lexical decision times, and self-paced listening times increase at points in a sentence where models of sentence processing predict an increased processing load (Caplan et al. 1994; Ferreira et al. 1996a; 1996b; Ford 1983; Frazier & Rayner 1982; King & Just 1991). Lexical decision reaction times to visually presented words and detection times for extraneous noises and phonemes increase at these points during the auditory presentation of a sentence (Cohen & Mehler 1996; Frauenfelder et al. 1980; Zurif et al. 1995), consistent with the view that local processing load is higher at these points. These and many more studies suggest that syntactic analyses and the use of these analyses to assign sentence meaning require processing resources.

As noted above, we contrast the process whereby a listener or reader extracts the meaning of the sentence from the signal with the process of using the meaning that has been extracted to perform other tasks. To appreciate the difference intuitively, consider sentence (3):

3. Please pick up four tomatoes, a pound of apricots, prune juice, shallots, six apples, and a bag of carrots on the way home.

In comparison to sentence (1), it is relatively easy to understand what sentence (3) means. However, certain operations that one could perform on the meaning of the sentence would be difficult, such as carrying out the request from memory. We draw a distinction between these types of operations, which include remembering the content of a sentence, using the meaning of a sentence to plan action, reasoning on the basis of sentence meaning, and other aspects of what we call post-interpretive processing, and sentence interpretation itself.

Utilizing the propositional content of a sentence to accomplish tasks usually involves controlled, conscious processing, as opposed to the largely unconscious processes involved in sentence interpretation. Controlled, conscious processes constitute the domain of verbally mediated tasks that is thought to require processing resources (Schneider & Shiffrin 1977; Shiffrin & Schneider 1977). There are many different purposes to which the propositional content of a sentence can be put, and it is widely thought that, to the extent to which these purposes involve controlled, conscious processing, they all make demands on verbal working memory (Baddeley 1986).

In many experiments on working memory, researchers have contrasted performance on pairs of sentences that are selected to vary in their complexity. In most of these experimental manipulations, there is some effect of the change in sentence type on both interpretive and post-interpretive processing. Some of these contrasts, however, increase the working memory load primarily at the interpretive stage of sentence processing, others at the post-interpretive stage. Two contrasts in particular will recur in the presentation of data to follow: that between sentences with subject versus object relatives (*The boy that hugged the girl kissed the baby; the boy that the girl hugged kissed the baby*) and that between sentences with one proposition and two propositions corresponding to each of two verbs (*The boy hugged the girl and the baby; the boy hugged the girl and kissed the baby*). Although both of these contrasts involve changes in both the syntactic structure of a sentence and aspects of its meaning, we will present evidence indicating that the first of these contrasts increases processing load primarily at the syntactic level and the second increases processing load primarily associated with different aspects of post-interpretive processing.

2. Sentence comprehension and working memory capacity in normal subjects

As noted above, our focus is on the nature of the working memory system involved in the interpretive aspect of sentence processing. One possibility, advanced by Just, Carpenter, and their colleagues, is that humans have a set of verbal processing resources that can be devoted to all verbal tasks (Just & Carpenter 1992; King & Just 1991; MacDonald et al. 1992; Miyake et al. 1994). An alternative view is that part of the verbal working memory system is specialized for interpretive aspects of sentence comprehension, specifically, assigning syntactic structure and using it to determine the meaning of a sentence (Caplan & Waters 1995; 1996; Waters & Caplan 1996a; 1996b; 1996c; Waters et al. 1995).

There have been two basic approaches to investigating the possible specialization of working memory for sentence interpretation. One has been to determine the relationship of individual differences in verbal working memory capacity to the efficiency of sentence interpretation. A second has been to investigate the pattern of mutual interference (or noninterference) of sentence interpretation and concurrently holding a verbal load in short-term memory.

With respect to the individual-differences approach, the single-resource (SR) theory predicts that having a low working memory capacity will reduce the resources available for sentence processing and make it less efficient; a separate-sentence-interpretation-resource (SSIR) theory predicts that performance on general verbal working memory tasks will not predict language processing efficiency. The SR model is therefore supported by significant correlations between measures of working memory capacity and measures of sentence processing efficiency. Many studies correlate working memory capacity and performance on verbally mediated tasks (see Daneman & Merikle, 1996, for review). However, almost no research into the role of working memory in sentence processing has used correlational analyses. Instead, research into this question that uses the individual-differences approach has been based on experiments in which performance on a working memory task serves to divide subjects into those with high and low (and sometimes medium) working memory capacity, and sentence processing performance is measured. The single-capacity model predicts that, in the absence of ceiling and floor effects, there will be a main effect of working memory capacity in experiments of this type, with high-capacity subjects performing better on the sentence processing task than low-capacity subjects. The SSIR theory might be thought to predict that there will be no main effect of group in such analyses. However, there is another issue that complicates this prediction: low-capacity subjects would not be expected to be as adept at accomplishing many psycholinguistic tasks as high-capacity subjects. For instance, King and Just (1991) had subjects carry out a self-paced reading task while simultaneously remembering the last word of each of a set of sentences (sets ranged from one to three sentences). Poorer performance by low-span subjects could be a reflection of difficulties they had with dividing their attention in this task. Even tasks that are relatively simple in experimental psycholinguistics, such as self-paced reading coupled with answering questions at the end of each sentence (MacDonald et al. 1992) or rapid serial visual presentation (RSVP) with end-of-sentence acceptability judgments (Waters & Caplan 1996b), might be harder for subjects with low working memory capacities. In general, poorer performance by low-capacity subjects is not necessarily due to their inability to accomplish the processing that is the focus of an experiment but might represent difficulties these subjects have with other aspects of task demands. Hence, group effects in sentence processing may be compatible with the SSIR theory under some circumstances.

In most experiments sentence materials have been constructed to represent two points along a complexity scale rather than a single sentence type. This is because most researchers have assumed that both low- and high-capacity subjects have sufficient resources to assign structure and meaning in simple sentences and that the limitations imposed by low working memory capacity are felt only or mostly with complex sentences (see, e.g., King & Just 1991; Miyake et al. 1994). Although there has been little discussion of this assumption in the literature, it seems at least plausible, given the number of sentences that most language users appear to process without difficulty. Given this assumption, the SR theory predicts that there will be an interaction between syntactic complexity and working memory capacity: sentence complexity will affect low-capacity subjects more than high-capacity subjects, and the differences between performance of low- and high-capacity subjects with complex sentences will be greater than the differences between the groups with simple sentences.[1] Even if the SSIR theory is compatible with a main effect of group in such experiments for the reason outlined above, it predicts that no such interactions will be found. An interaction between sentence complexity and working memory group therefore favors the SR theory.[2]

A second approach is to investigate the pattern of mutual interference (or noninterference) of verbally mediated tasks. According to the SR model, verbal memory loads that are imposed external to the comprehension task (such as a concurrent digit span task) and sentence comprehension draw on the same pool of resources. According to SSIR theory, "interpretation-external" and "interpretation-internal" processes draw on different resource pools. As with the relationship between working memory capacity and sentence processing, research using this approach has assumed that a concurrent load will affect the processing of simple sentences less than that of complex sentences (Baddeley & Hitch 1974; Miyake et al. 1994). With this assumption, SR theory predicts an interaction between external load and sentence complexity, and SSIR theory does not. In research using this approach, the concurrent task must utilize central executive resources. Digit span has been the most widely used task for this purpose.[3]

These two approaches can be combined. In SR theory, comprehending more complex sentences and maintaining a larger load in verbal memory both require more processing resources from the same pool, and this pool is smaller in some subjects. Thus, SR theory would be supported by the finding that the expected impairment of low-capacity subjects on more complex sentences is exacerbated by a concurrent memory load. In contrast, SSIR theory holds that maintaining a digit load in memory calls on a resource pool that is standardly measured by working memory tests and that comprehending sentences calls on another pool of resources. This model therefore predicts that low-capacity subjects should perform less well than high-capacity subjects under conditions of increased verbal memory load, but it does not predict that this effect will be greater for syntactically complex sentences.[4]

We review here the evidence relevant to each of these lines of argument. For published papers that include a large number of analyses, we present summaries of the results. For work in our laboratory that has not yet been published, we present results and illustrations of representative and critical findings.

2.1. Individual differences in speed and accuracy of syntactic processing. Relating individual differences in verbal working memory capacity to syntactic processing abilities requires that we be able to measure each individual's working memory capacity. Daneman and Carpenter (1980) developed a "reading span" task that has become the standard method of assessing verbal working memory. In this task, subjects are required to read aloud increasingly longer sequences of sentences and to recall the final word of all the sentences in each sequence. A subject's working memory capacity is defined as the longest list length at which he or she is able to recall the sentence-final words on the majority of trials. There are various versions of this test (Tirre & Pena 1992; Turner & Engle 1987; Waters et al. 1987) as well as other tests with quite different sets of operational demands, such as random digit generation (Petrides et al. 1993) or backwards digit span (Kemper 1988), that have been used to measure verbal working memory. Most research into the performance of groups who differ in working memory capacity in processing sentences with different structures has used the Daneman and Carpenter task as the measure of working memory, and the results that we will report follow this practice.

The first set of results deals with interpreting sentences with relative clauses. There is a considerable amount of evidence that assigning the structure of an object-relativized sentence such as sentence (4) is more demanding than a subject-relativized sentence such as sentence (5), and this increased load occurs at the verb of the embedded clause (*push*) and/or the main verb (*kiss*).

4. The boy that the girl pushed kissed the baby.
5. The girl that pushed the boy kissed the baby.

King and Just (1991, exp. 1) reported self-paced word-by-word reading times for high- and low-span subjects for these two sentence types. In this task, a sequence of words that forms a sentence is presented visually, one word at a time, with the subject pressing a response key to see each successive word. Reading time increases at points of syntactic complexity in this task. King and Just (1991) present a graph, reprinted by Just and Carpenter (1992), that is said to show that the biggest reading time differences between high- and low-span subjects were in the syntactically critical area of the object-relative sentences. However, no statistical analyses were reported by King and Just (1991) to support the contention that the difference in reading times between high- and low-span subjects is specifically localized to the region of object-relative sentences where there is the greatest processing load. Moreover, the data presented by Just and Carpenter (1992) and by King and Just (1991) did not isolate performance on sentences in which the subjects did not have to retain the sentence-final words (no analyses were reported by King and Just on the sentences in the zero-load condition alone).

Our laboratory has failed to find differences between high- and low-span subjects in on-line processing of similar types of sentences. In one study, we used the "auditory moving windows" task, in which spoken sentences are recorded, digitized, and segmented into constituents, and subjects press a key to hear each successive constituent (Ferreira et al. 1996b). As in the study by Ferreira et al. (1996a), we found significant increases in listening times for the embedded verb of object-relativized sentences compared to subject-relativized sentences. However, in a group of 100 subjects, this effect was no larger in low-capacity than in high-capacity subjects. The results are displayed in Figure 1, which shows the differences in listening time for the same phrases in the complex object-relativized sentences and the simple subject-relativized sentences. As Figure 1 shows, listening times increase at the embedded verb and, to a lesser degree, at the main verb – both points of syntactic processing load – and these increases do not differ among the three groups of subjects. We have also used a continuous visual lexical decision task (Ford 1983), in which subjects were required to judge whether each successive word of the sentence was a real word. We replicated Ford's finding of a statistically significant increase in reading time on the relative clause verb in object- as compared to subject-relative clauses. In a group of 98 undergraduates

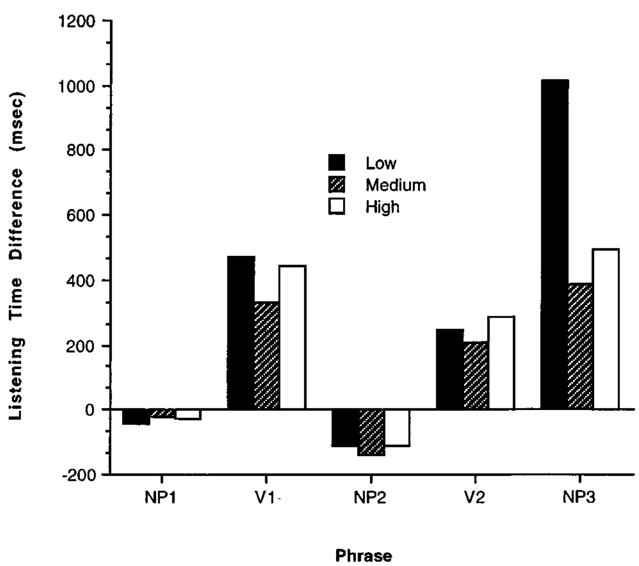


Figure 1. Differences in self-paced listening times for each phrase for object- minus subject-relative sentences (The boy that the girl pushed kissed the baby; The boy pushed the girl that kissed the baby) in subjects with low, medium, and high working memory capacity. NP1, first noun phrase; V1, first verb; NP2, noun in relative clause; V2, second verb; NP3, sentence-final noun. High, high-capacity subjects; medium, medium-capacity subjects; low, low-capacity subjects.

divided into high-, medium-, and low-span subjects using the Daneman and Carpenter task, this effect was no larger in low-span than in high-span subjects. We have obtained similar results in a replication of this study with 63 subjects in which the sentences were presented auditorily rather than visually.

To summarize these data, we know of four studies that compare the performance of high- and low-capacity subjects on measures of word processing as a function of sentence type (object- vs. subject-relativization) and region within sentence: the self-paced reading study of King and Just (1991) in which subjects had to recall sentence-final words as well as process the sentences, our self-paced auditory and visual lexical decision studies, and our self-paced listening (auditory moving window) study. None of these studies showed an interaction between span group, sentence type, and region. In our studies, there were no interactions of span group and sentence type in the region at which the greatest processing load occurs.

A second type of sentence that has heavy processing demands is a so-called "garden path" sentence. These sentences are ones that are locally ambiguous, and that eventually are resolved in favor of the less-preferred interpretation. For instance, the sequence *The experienced soldiers warned about the dangers* is ambiguous. The phrase *warned about the dangers* could be a main verb with a prepositional phrase adjunct, as in (6), or a relative clause in which the relative pronoun has been omitted, as in the garden path sentence (7).

6. The experienced soldiers warned about the dangers before the midnight raid.

7. The experienced soldiers warned about the dangers conducted the midnight raid.

MacDonald et al. (1992) claim to have found differences in high- and low-capacity subjects' self-paced reading times and accuracy in answering questions about sentences such as sentence (7). MacDonald et al.'s results are very complex, and the reader is referred to Waters and Caplan (1996a) for a review and detailed analysis. In general, for garden path sentences such as sentence (7), group differences in reading times and accuracies were not statistically significant and differences in reading times were not found while subjects were reading the ambiguous portions of the sentence.

We have examined the performance of high- and low-capacity subjects on sentences such as sentence (7), as well as two other types of garden path sentences (Waters & Caplan 1996b). We carried out three experiments in which we compared the ability of high- and low-span subjects to comprehend these three different types of garden path sentences both with visual presentation of whole sentences and with RSVP at increasingly faster rates. Although garden path sentences were significantly more difficult to interpret than nonambiguous control sentences, these effects were not larger in subjects with low working memory capacity. We also replicated the MacDonald et al. (1992) self-paced experiment using their materials in 91 subjects and failed to find memory span group differences in reading times (Fig. 2), a replication of MacDonald et al.'s result, which, as noted above, did not find significant group differences in reading times in the ambiguous region of these sentences.

MacDonald et al. (1992) also examined self-paced reading times for sentences such as sentence (6), which are locally ambiguous but are resolved in favor of the preferred interpretation and the simpler syntactic structure. They

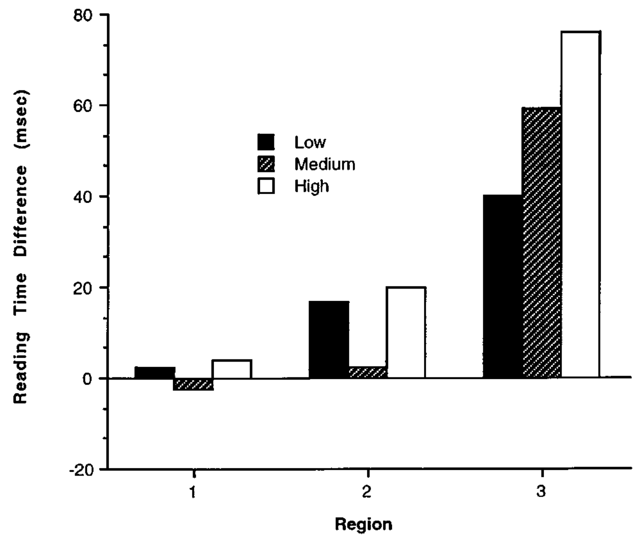


Figure 2. Differences in self-paced reading times for garden path minus control (unambiguous) sentences in subjects with low, medium and high working memory capacity as a function of region of the sentence. Region 1 contains the ambiguous verb and associated prepositional phrase (*warned about the dangers*). Region 2 contains the disambiguating segment (*conducted the midnight*). Region 3 contains the end of the sentence (*raid*). High, high-capacity subjects; medium, medium-capacity subjects; low, low-capacity subjects.

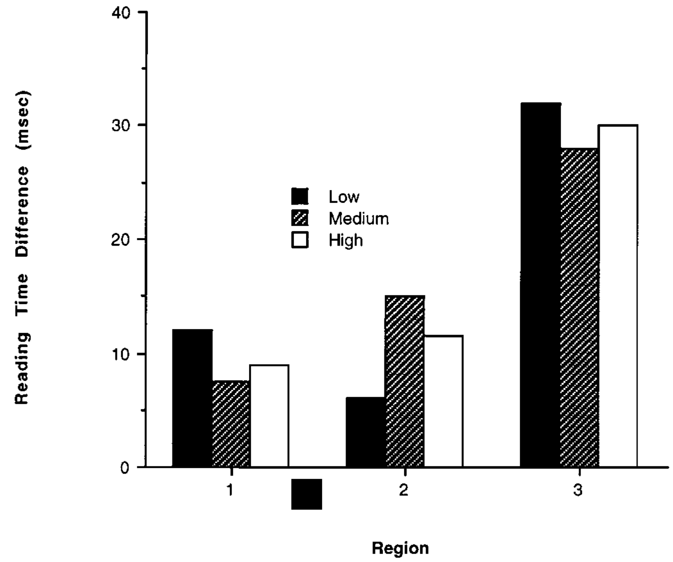


Figure 3. Differences in self-paced reading times for preferred (main verb) interpretation of ambiguous minus unambiguous sentences in subjects with low, medium, and high working memory capacity as a function of region of the sentence. Region 1 contains the ambiguous verb and associated prepositional phrase (*warned about the dangers*). Region 2 contains the next segment (*before the midnight*). Region 3 contains the end of the sentence (*raid*). High, high-capacity subjects; medium, medium-capacity subjects; low, low-capacity subjects.

found that high-capacity subjects had longer reading times than low-capacity subjects, and they accounted for this superficially paradoxical result by suggesting that high-capacity subjects maintain both interpretations of a sentence in mind until the ambiguity is resolved, whereas low-capacity subjects lack the resources to do so and drop the less preferred analysis. However, their data provide little evidence to support the view that subjects with different working memory capacities perform differently in processing these sentences. MacDonald et al. did find that reading times were longer in high-capacity than in low-capacity subjects, but this was true only for the last word of sentences such as sentence (6) and not throughout the ambiguous region. Moreover, low-capacity subjects made more errors in answering questions. This pattern may simply represent different speed–accuracy trade-offs in the decision-making process in the different capacity groups. Pearlmutter and MacDonald (1995) reported a similar result, in which high- and low-capacity subjects used different cues regarding the likelihood that the sentence would end with a main clause or a reduced relative clause interpretation. All of MacDonald and her colleagues' results must be interpreted cautiously because she had subjects answer questions about the thematic role played by the first noun after each sentence. This may have alerted the more verbally adept high-span subjects to the presence of the ambiguity, and any differences between groups may be due to strategic factors operating in these experiments.

We have repeated MacDonald et al.'s experiment in 91 subjects using the exact methods and materials from MacDonald et al. (1992). We did not find poorer performance in high- than in low-capacity subjects on sentences such as sentence (6) (Fig. 3). In a follow-up experiment, we eliminated the questions that were posed after the sentences. This experiment, with 75 subjects, also failed to find differences between high- and low-capacity subjects.

A third line of research has investigated the interaction of semantic and syntactic information in sentence interpretation in subjects with different working memory capacities. Just and Carpenter (1992) studied garden path sentences similar to sentence (7) but in which the animacy of the first noun constrains the possible interpretation of the sentence, as well as sentences that were unambiguous because of the presence of a relative pronoun. The sentence types are shown below:

8. *Syntactic garden path; semantically unconstrained:* The defendant examined by the lawyer shocked the jury.

9. *Syntactically unambiguous; semantically unconstrained:* The defendant that was examined by the lawyer shocked the jury.

10. *Syntactic garden path; semantically constrained:* The evidence examined by the lawyer shocked the jury.

11. *Syntactically unambiguous; semantically constrained:* The evidence that was examined by the lawyer shocked the jury.

Their data show that both high- and low-capacity subjects have longer first-pass fixation durations on the *by* phrase in the sentences with reduced relative clauses (sentences 8 and 10) than in the corresponding sentences with the overt relative clause marker (sentences 9 and 11), regardless of the animacy of the first noun. This provides evidence for similar degrees of modularity of syntactic processing in both high- and low-span subjects (Ferreira 1992; Waters & Caplan 1996a).

Just and Carpenter also found that the high-capacity subjects showed longer fixation times on the *by* phrase in rela-

https://doi.org/10.1017/S0140525X99321786 Published online by Cambridge University Press

tive clause sentences with animate first nouns (sentences 8 and 9) than in the corresponding relative clause sentences with inanimate first nouns (sentences 10 and 11), whereas this was not true for the low-capacity subjects. This result is difficult to interpret. It indicates that high-capacity subjects were unable to use the clear, unambiguous cue to syntactic structure afforded by the relative pronoun and the auxiliary verb (*that was*) of the relative clause in sentence (9), and took *the defendant* as the subject of the active form of the verb *examined* in the phrase *The defendant that was examined,* because of the animacy of the word *defendant.* It is hard to understand why the more verbally adept high-capacity subjects would not use this clear cue, and other data show the opposite pattern. King and Just (1991, exp. 2) studied the ability of high- and low-capacity subjects to comprehend object-relative sentences that contained verbs that either did or did not provide strong pragmatic cues regarding which of the two potential actors in the sentence was the agent of a given verb (sentences 12–15):

12. *Main verb constrained; embedded verb constrained:* The robber that the fireman rescued stole the jewelry.
13. *Main verb unconstrained; embedded verb constrained:* The robber that the fireman rescued watched the program.
14. *Main verb constrained; embedded verb unconstrained:* The robber that the fireman detested stole the jewelry.
15. *Main verb unconstrained; embedded verb unconstrained:* The robber that the fireman detested watched the program.

King and Just found that the comprehension of high-capacity subjects did not improve in the pragmatic bias condition, but that of low-capacity subjects did. Although these findings reflect accuracy on end-of-sentence judgments rather than eye-fixation durations during on-line sentence processing, they are the opposite of the Just and Carpenter (1992) results.

A fourth line of research that has been pursued involves increasing the processing demands associated with sentence comprehension by changing the perceptual demands of the task. Miyake et al. (1994) reported a series of experiments using RSVP, in which low-, medium-, and high-capacity subjects were required to indicate the actor or answer questions about sentences that required syntactic analyses to be understood. Miyake et al. reported significant interactions between span group and sentence type. To determine whether these interactions reflected differences in the ability of subjects with different working memory capacities to utilize syntactic structures, we reanalyzed the data from the one experiment for which sufficient data were available for reanalysis and found that subjects with different working memory capacities did not perform differently as a function of syntactic complexity (Caplan & Waters 1995). The group-by-sentence-type interaction was due to poorer performance by low-span subjects on sentences with two propositions compared to sentences with one proposition. The larger effect of the number of propositions in low-capacity subjects suggests that these subjects have more difficulty than high-capacity subjects retaining in memory information about the actor in each proposition. Thus, this study provides evidence that increasing the perceptual demands of a task does not differentially affect the abilities of subjects with different working memory capacities to process syntactic structure but may make certain post-interpretive processes more difficult for low-capacity subjects.

We conclude this section with a brief discussion of the effects of age on sentence processing. Most studies have found that working memory declines with age (Salthouse 1991), so the study of age-related changes in sentence comprehension provides indirect evidence about the effect of changes in working memory capacity on this function.

The effects of aging on sentence comprehension abilities have been examined using a variety of tasks such as object manipulation (Feier & Gerstman 1980), question answering (Davis & Ball 1989; Emery 1985), acceptability judgment (Kemper 1988; Obler et al. 1991), and sentence recall (Norman et al. 1991). Some studies have found that elderly subjects perform more poorly overall, but that they are not differentially impaired on syntactically more complex sentences (e.g., Feier & Gerstman 1980). Several studies have measured sentence processing as a function of age using on-line tasks. Baum (1991) and Waldstein and Baum (1992) used a word-monitoring paradigm, in which subjects were required to monitor for a word that followed an ungrammaticality in a sentence, to investigate the sensitivity of younger and older adults to local and long-distance ungrammaticalities. Although processing times and error rates of the elderly were higher overall than those of the younger subjects, there was no evidence that elderly subjects were more reliant on sentential context or that they were less sensitive to ungrammaticalities. Kemtes and Kemper (1996) also found that the effect of syntactic ambiguity on word-by-word reading times did not differ in older and younger subjects, although the older subjects were more affected on an off-line measure of accuracy in responding to questions. We analyzed the data from the garden path experiment described above (Waters & Caplan 1996b) for effects of age, and found no such effects.

Other studies have shown decreased comprehension of more complex syntactic structures among the elderly (e.g., Davis & Ball 1989; Obler et al. 1991). However, in terms of the distinction we made above between interpretive and post-interpretive processing, the tasks on which aging subjects have shown such effects are ones that require elaborate post-interpretive processing, such as retaining and reordering large amounts of material in memory (Kemper 1986; Light 1990) or interpreting implausible sentences (Davis & Ball 1989; Obler et al. 1991). Thus, in many of these studies, the effect of age may be attributed to difficulties subjects might have with processes other than those involved in sentence interpretation. In the domain of on-line processing, using a cross-modal naming paradigm, Zurif et al. (1995) found that older subjects were delayed in establishing the connection between a gap in a relative clause and the head of the clause (e.g., determining that it is *the man* who is kissed in a sentence such as *The man who the woman kissed was embarrassed*). Zurif et al.'s data are difficult to interpret, however, insofar as a comparison group of younger subjects was not tested on the same materials. We have failed to find differences as a function of age in listening times for phrases in high-load positions in more complex sentences in the auditory moving window task described above. Overall, the evidence indicates that there is little, if any, effect of aging on syntactic processing in comprehension in aging and, therefore, indirectly supports the view that differences in working memory capacity are not associated with differences in the efficiency of syntactic processing in sentence comprehension.

In summary, several studies have examined the performances of subjects with different verbal working memory capacities in syntactic processing in sentence comprehension tasks. Most of these studies have not shown differences as a function of span group, even in processing sentences with very complex syntactic structures. A few studies have reported group differences, but these differences might be due to strategic effects (e.g., different speed–accuracy trade-offs in different capacity groups) and are inconsistent across a wider set of studies. Overall, this literature provides evidence that the speed and accuracy of syntactic processing does not differ in a systematic way in normal subjects with different working memory capacities as measured by performance on sentence span tasks. In contrast, there is a suggestion in the data that low-capacity subjects are more affected than high-capacity subjects by some sentential features, in particular, the number of propositions in a sentence. Given that subjects with different working memory capacities do not differ in their ability to use syntactic structure to determine sentence meaning, the larger effect of the number of propositions among low- versus high-capacity subjects suggests that low-capacity subjects have more difficulty with retaining in memory information about the propositional content of a sentence. We shall present data pertaining to other post-interpretive functions below.

2.2. Effects of an external memory load on speed and accuracy of syntactic processing. The second approach to the issue of the possible specialization of working memory for syntactic processing is to investigate the pattern of mutual interference of verbally mediated tasks. According to the SR model, verbal memory loads that are imposed external to the comprehension task (such as a concurrent digit span task) and sentence interpretation draw on the same pool of resources, and the SR model therefore predicts that there will be interference between the two. In SSIR theory, "interpretation-external" and "interpretation-internal" factors draw on different resource pools; hence the theory predicts main effects of each but no interference between the two. We have undertaken several studies; all found the second pattern.

In the first study (Waters et al. 1995), we had subjects match sentences to pictures in a no-interference condition and in two concurrent verbal load conditions: while retaining a random sequence of digits equal to their span or equal to one less than their span. Sentences were all semantically reversible and varied in their syntactic complexity. The effect of syntactic complexity was examined by comparing performance on pairs of sentences that were matched for number of words, propositions, verbs, and thematic roles but that differed with respect to whether a sentence contained a noun phrase that had been moved, according to Chomsky's (1986; 1993) theory, creating a nonstandard order of thematic roles. For instance, object-relativized sentences (e.g., *The boy that the girl pushed kissed the baby*) were syntactically complex compared to subject-relativized sentences (e.g., *The girl pushed the boy that kissed the baby*) according to this measure. There was an effect of load on the sentence accuracy scores but no effect of syntactic complexity and no interaction of syntactic complexity with load in these scores or in the digit-recall data. We reexamined this question by changing the task to enactment, which may be more demanding (Waters & Caplan, submitted). Subjects had to enact the thematic roles in sentences with the same structures used in the picture-matching task, in a no-interference and a digit-load condition. Again, there was an effect of condition but no interaction of load and syntactic complexity.

These studies used off-line measures of accuracy to assess sentence comprehension, which may not be sensitive to the effects of a concurrent load. We have carried out one dual-task study in which subjects' performances were timed as well as scored for accuracy (Waters et al. 1987). In this study, subjects made judgments about the plausibility of sentences that differed in their syntactic complexity (object vs. subject relativization) in no-interference and digit-load conditions (recalling a random sequence of six digits). There were effects of syntactic type and of load; the interaction of load and syntactic complexity was significant in the sentence data (RTs and accuracy) in the analyses by subjects but not by items.

Although the above results suggest that digit span does not interfere with syntactic processing in sentence comprehension, other studies indicate that digit span does interfere with several verbally mediated tasks (Baddeley & Hitch 1974), and several researchers have interpreted their results as showing that digit span affects subjects' abilities to process syntactically complex sentences. However, a review of this literature indicates that the effect of a digit load on syntactic processing occurs only under some task conditions. We summarize this literature here; more complete discussions can be found in Caplan and Waters (1996) and Waters et al. (1995).

Baddeley and Hitch (1974) found that a concurrent random six-digit load interfered more with subjects' performance on passive than on active sentences in a "reasoning task" in which subjects had to indicate whether a statement about a sequence of letters corresponded to the actual presentation of letters (e.g., *A is not followed by B – BA*). However, many factors (order of mention of the letters and the letter pair, presentation of letters in their alphabetical order in either the proposition or the pair, proactive interference, etc.) that might have affected the results were not considered in the analyses, and the interaction of load and voice may have been qualified by higher order interactions that were not considered in this early work on this topic. Several investigators (Gick et al. 1988; Morris et al. 1988; 1990; Wright 1981) have studied younger and older subjects on tasks that combine digit load with a sentence-plausibility (or truth)-judgment task. Many of these studies have found significant two-way interactions between the size of a concurrent memory load and sentence complexity. The sentence complexity metric in these experiments, however, was the presence of a negative element in the sentence (e.g., *A cat does not hunt mice)*, which significantly affects the complexity of the judgment task but has little effect on the complexity of the syntactic structure of the sentence. These studies have not shown interactions of syntactic structure with load.

Other experiments have investigated the effect of a concurrent digit load that has been presented as the requirement to recall the final word of each of a series of sentences. King and Just (1991) had subjects read object- and subject-relative sentences word by word in a self-paced task, while retaining one, two, or three sentence-final words in memory, and answer questions about the sentences. There was an interaction between syntactic complexity and memory load in the recall data. There was no such interaction in the accuracy data, and the reading time data were not analyzed for the effect of size of memory load. A second experiment using the Daneman and Carpenter paradigm was reported by Carpenter and Just (1988). In this experiment, sentences were considered more difficult if they contained less frequent and less concrete words, and so the effect of a memory load on syntactic processing was not investigated. Results using a variant of the Daneman and Carpenter paradigm were reported by our group (Waters et al. 1987; Waters & Caplan 1996b). We found an effect of syntactic complexity (object vs. subject relativization) on the number of sentence-final words recalled in a sentence-acceptability-judgment task. Finally, Wanner and Maratsos (1978) interrupted sentence presentation to allow a set of words to be displayed and reported poorer recall of words in object relativized than in the less complex, subject-relativized sentences.

This pattern of results suggests that whether or not a concurrent digit load interferes with syntactic processing depends on the relationship of the recall and sentence processing tasks. The published literature is consistent with the generalization that when the stimuli in both the sentence and the recall tasks are presented in an uninterrupted fashion – as when the digit load is presented prior to each sentence and must be recalled once the sentence has been understood – no interference is found, but when the presentation of the stimuli in one task is interrupted by the presentation of a stimulus relevant to the second – as when the sentence is interrupted by a series of words while it is being presented, or the presentation of the items to be recalled is staggered across the sentence task (as in sentence-final word recall) – a digit load of sufficient size interferes with processing syntactically complex sentences more than with processing syntactically simple sentences. This pattern of results suggests that the reason for the effect of a digit load on syntactically complex sentences is not that retaining a sequence of digits and structuring a sentence syntactically compete for the same resource pool but, rather, that the attentional shifts associated with interrupting each task interfere with subjects' abilities to structure sentences syntactically or to use that structure to assign sentence meaning. We have suggested that this might be a secondary effect of disruptions of lexical access (Caplan & Waters 1996). (An anonymous *BBS* referee pointed out that interruption of a task makes it harder to retrieve information associated with that task, which also might affect the ability to construct more complex structures.)

An important point to note is that the tasks in which an interaction between digit load and syntactic complexity has *not* been found are not simply so easy that this interaction cannot be detected. Testing for digit recall at each subject's predetermined span guarantees that each subject will be below ceiling on the recall task. In fact, in the research cited above, the percentage of trials in which all digits were produced in correct serial position in the recall task was usually about 50%. The subjects who have been tested in the literature cited above also performed below ceiling on at least some sentence types in which the effects of syntactic complexity could be examined. Nonetheless, a concurrent digit load did not lead to an effect of syntactic complexity except when the presentation of the sentence or the digit sequence was interrupted by the other stimulus.

In contrast to the failure of a concurrent digit load to affect processing of syntactically more complex sentences (except when either stimulus is interrupted), the same external verbal memory load has significant effects on sentences with two propositions compared to sentences with one. Waters et al. (1987) found that a concurrent load of six random digits interfered more with acceptability judgments about sentences with two propositions than sentences with one. Waters et al. (1995) reported that this was also the case in a sentence-picture-matching task when the concurrent digit load was equal to each subject's span, and Waters and Caplan (submitted) have found similar results using an enactment task. This suggests that, unlike the case with syntactic processing in sentence comprehension, operations on the propositional content of a sentence such as matching it to knowledge in semantic memory or depictions of events or using it to plan and execute actions share resources with span tasks.

2.3. Effect of the combination of external memory load and individual differences in working memory capacity on speed and accuracy of syntactic processing. A third approach to the question of the specialization of verbal processing resources is the combination of the first two approaches, looking for the combined effects of individual differences in working memory capacity and of external load on language processing efficiency. The SR theory asserts that comprehending more complex sentences and maintaining a larger digit load in memory both require more processing resources from the same pool, and that this pool is smaller in some subjects. Thus it predicts that the expected impairment of low-capacity subjects on more complex sentences will be exacerbated by a concurrent memory load. In contrast, the SSIR theory holds that maintaining a digit load in memory calls on a resource pool that is standardly measured by tests such as the Daneman and Carpenter reading span task and that comprehending sentences calls on another pool of resources. Therefore, this model predicts that low-capacity subjects should perform less well than high-capacity subjects under conditions of increased verbal memory load, but it does not predict that this effect should be greater for syntactically complex sentences.

Relevant data come from several studies. King and Just (1991) had high-, medium-, and low-capacity subjects read object- and subject-relative sentences word by word in a self-paced task while retaining one, two, or three sentence-final words in memory. For the recall task, King and Just reported significant interactions between group and size of memory load and between sentence type and size of memory load but no interaction between group, syntactic complexity, and memory load. In the probe question comprehension results, there was a statistically significant main effect of group, and the effect of sentence type was marginally significant, but none of the interactions involving memory load approached statistical significance. The analyses reported by King and Just (1991) thus do not show the interactions between group, sentence type, and load that the SR theory predicts.

Carpenter and Just (1988) investigated the number of words that high-, medium-, and low-capacity subjects recalled in a Daneman and Carpenter-type reading span task when the sentences were either "easy" or "difficult." All subjects performed the task at set sizes 2, 3, and 4 under conditions in which the stimulus materials contained 0, 1, or all difficult sentences. High-capacity subjects were influenced by difficult sentences only at set size 4, whereas medium-capacity subjects were affected by difficult sentences at set sizes 3 and 4. However, the authors did not report on the presence or absence of a three-way interaction between span group, set size, and sentence type, so it is unclear whether high- and medium-capacity subjects were differentially affected at either set size 3 or set size 4 by sentence complexity. Contrary to the predictions of the capacity theory, the performance of low-capacity subjects was unaffected by sentence difficulty at any set size. In addition, in this experiment, sentences were considered more difficult if they contained less frequent and less concrete words, so the experiment does not bear directly on syntactic processing.

We investigated the possibility that syntactic processing would be more affected by a concurrent verbal memory load in low-capacity than in high-capacity subjects in enactment (Waters & Caplan, submitted), sentence-picture-matching, and speeded sentence-acceptability-judgment tasks. Figure 4 shows the data from the enactment task. The effect of syntactic complexity was greater in the low- than in the medium-capacity group, but did not differ in the low- and high-capacity groups. This effect did not increase more in the low- than in the medium- or high-capacity subjects in the dual-task conditions. Similar results showing no increase in the magnitude of a syntactic complexity effect in low-capacity subjects compared to high-capacity subjects as a function of the presence of a digit load were found in the other tasks.

2.4. Some considerations regarding studies of normal subjects. The results of the studies reviewed above are consistent with the SSIR model. However, they are subject to several caveats, which we will discuss here.

The first is that many of the arguments made in the preceding review are based on null results – the failure to find differences between different capacity groups in syntactic processing, to find effects of a concurrent verbal memory load on syntactic processing, or to find differential effects of load on syntactic processing in low- versus high-capacity subjects. Issues of power and related concerns arise in interpreting nonsignificant results. Concerns about the interpretation of null results can be completely laid to rest only

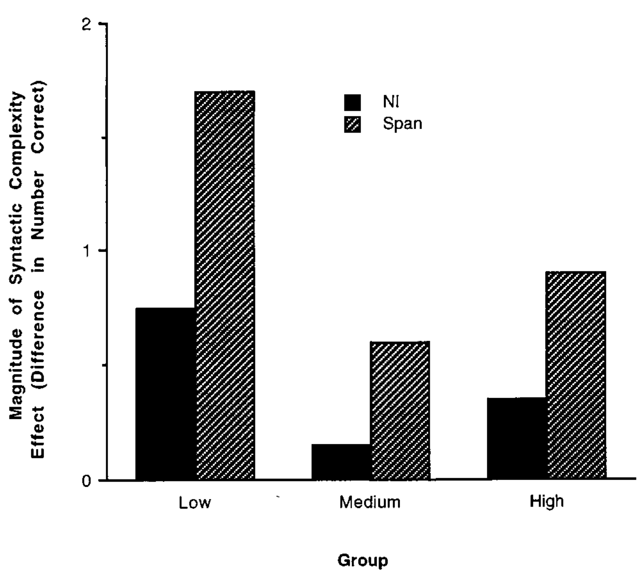


Figure 4. Performance of high-, medium-, and low-span subjects on an enactment task with and without a concurrent digit load. NI, no interference; Span, concurrent span condition.

by running a much larger number of subjects than is feasible (often hundreds or more; Krueger 1994). However, they can be partially countered by noting there were interactions between sentence type and condition and sentence type and group in many of the experiments we have reported, but that the sentence type variable entering into these interactions was one that reflected not syntactic complexity but rather the number of propositions in a sentence. This suggests that the designs had sufficient power to generate interactions involving the sentence type factor but that varying the syntactic form of a sentence does not differentially affect low-capacity subjects or performance under dual-task conditions.

Other considerations pertain to the dual-task experiments. One issue is that most work has used the digit span task as a concurrent verbal memory load designed to compete for resources with sentence comprehension. The span task has the advantages that load can be equated across subjects by testing each subject at his or her span and that span represents a patient's ceiling and thus calls on all resources that can be devoted to immediate serial recall. Nonetheless, span may be largely based on a specialized auditory-to-articulatory information transfer mechanism (Freidrich 1990), and additional research using concurrent tasks such as random number generation, which Baddeley (1993) has argued is less highly automatized than span and thus requires more of the limited resources of working memory, is needed to investigate the effect of a concurrent memory load on comprehension of syntactically more complex sentences. Preliminary results in our laboratory indicate that this task also does not interact with syntactic complexity. A second issue that arises regarding our interpretation of the dual-task experiments is that we have argued that the effects of a concurrent memory span task on syntactic processing are limited to situations in which the presentation of the stimuli relevant to one task is interrupted by the presentation of the stimuli relevant to the second. This suggests that the interference is secondary to switching attention, retrieving information when a task is interrupted, or other control processes, rather than shared resources. These possible explanations of the data must be tested directly.

Finally, our review of the literature on individual differences in working memory capacity and concurrent verbal memory loads on syntactic processing in sentence comprehension reveals that many findings with normal subjects have been hard to replicate. One possible reason for this difficulty is related to the measurement of subjects' working memory capacity. The most widely used measures of verbal working memory are sentence span tasks, based on the Daneman and Carpenter (1980) reading span. These tasks have several characteristics that affect how they may be interpreted. One is that, although these tasks are designed to require both processing and storage of verbal material, only the latter is usually measured. When measurements are made of performance on the sentence-processing component of a sentence span task, the measurement is usually only of accuracy, not of reaction time, and this measurement does not enter into the determination of which span group a subject belongs to. Because of this, subjects' spans may in part reflect any number of trade-offs between allocating attention and processing resources to the sentence and recall components of the sentence span task. A second consideration is that the verbal memory load in a sentence span task is unrelated to the sentence processing that is required in these tasks. This introduces a dual-task element into these tasks and allows them to be heavily affected by the capacity for dividing attention. Other working memory tasks (self-ordered number generation, in which subjects have to produce a random series of the digits 0–9 without repeating a digit, and externally ordered number generation, in which subjects have to recognize which digit is omitted in a random series of the digits 0–9; Petrides et al. 1993) require the memory and processing components of the task to be related to each other, as they are in naturally occurring tasks that require working memory. For all these reasons, performance on a sentence span task may not be a reliable assessment of working memory capacity.

We have investigated the relationship between different working memory measures and the reliability of several of these measures (Waters & Caplan 1996c). We tested 94 subjects on the Daneman and Carpenter (1980) reading span test, a variant of that test we developed in which RTs and accuracies on a sentence-acceptability-judgment task as well as sentence-final word recall were measured (Waters et al. 1987), and self-ordered and externally ordered number- and design-generation tasks (Petrides et al. 1993). There were significant correlations between all of these measures of working memory capacity (r between .52 and .58), other than the externally ordered design generation task. This level of correlation is consistent with other reports in the literature (Daneman & Merikle 1996; Engle 1995). However, test-retest reliability was not high. Forty-four subjects were retested on a subset of these measures at an average 1-month interval. Sentence-final word recall scores at the two testing sessions were weakly correlated ($r = .41$ for the Daneman & Carpenter task). Among the 44 subjects who participated in the follow-up study, 18 subjects (41%) changed in terms of their classification as high-, medium-, or low-capacity subjects at time 2, with equal numbers of subjects improving and declining and 22% of the subjects originally classified as high capacity being reclassified as low capacity on retesting. These figures call into question the stability of the most commonly used measures of working memory capacity. Difficulties in replicating effects of working memory on syntactic processing might thus reflect the insecurity of subject classification.

One way to deal with this problem is to use measures of working memory capacity that are more stable over time. We found that test-retest reliability improved when we used a composite measure of working memory capacity that took into account performance on the sentence processing as well as the recall portion of a sentence span test ($r = .88$; Waters & Caplan 1996c). However, little research has been carried out in which this type of measure is used as the basis for subject grouping. Another approach is to study patients with CNS disease as members of a low-capacity group because their performances are characteristically stable. There are a few studies that use this approach, to which we turn below.

3. Studies of patients with reduced working memory capacity

3.1. Patients with short-term memory disorders. One group of patients whose sentence comprehension has been studied fairly extensively are those with intact long-term memory but specific auditory verbal short-term memory impairments. Most of these patients have had problems with rehearsal and/or passive storage functions (see Vallar & Shallice, 1990, for representative cases). If working memory capacity plays a central role in language comprehension, then one might expect these patients' extreme limitations in the mechanisms that support central executive functions to have important effects on their language-comprehension abilities. The literature clearly establishes that this is not the case. In a review of all short-term memory patients published up to 1990, we found no evidence linking short-term memory impairment to difficulties with syntactic processing in sentence comprehension (Caplan & Waters 1990). The literature since that time bears out this pattern: although both storage and rehearsal impairments can affect certain aspects of comprehension (Martin & Feher 1990), these components of short-term memory have not been related to syntactically based sentence-comprehension difficulties. Results that are typical of the dissociation between impaired STM and intact comprehension were found in a patient, B.O., whom we tested quite extensively (Waters et al. 1991). This patient had a memory span of only two or three items when tested both on recall and probe-recognition tasks. When tested on the Daneman and Carpenter task, she had a working memory span of only one, despite the fact that she had no difficulty understanding the sentences used in this task when they were presented in isolation. B.O.'s performance on visual and auditory enactment tasks and a speeded visual whole-sentence-acceptability judgment task that tested comprehension of many complex syntactic forms was similar to that of normal individuals and much better than aphasic patients who have had a left hemisphere stroke. We have also tested B.O.'s comprehension of garden path sentences under timed auditory and written whole-sentence presentation conditions. She performed as well as normal individuals on these syntactically complex materials (Fig. 5). Parenthetically, B.O. showed a reliably larger effect of the number of propositions on comprehension than controls, in keeping with the idea that using the propositional content of a sentence to accomplish a task such as making a judgment about the plausibility of the sentence makes use of a short-term memory system that is not required for determining the sentence's syntactic form and propositional content.

However, as noted, the majority of patients with STM impairments have impairments in rehearsal and/or storage functions. The working memory system used in language comprehension has been claimed to correspond more closely to the central executive of a Baddeley-type model. The comprehension abilities of patients with limitations in the central executive component of this system thus constitute more direct and critical tests of the relationship between general-purpose working memory and comprehension.

3.2. Patients with limitations in the central executive. Patients with dementia of the Alzheimer type (DAT) have been found to have intact functioning of the articulatory rehearsal and phonological storage components of working memory but impairments on tasks that require central executive functions (Baddeley et al. 1991; Morris 1984; 1986; 1987; Morris & Baddeley 1988; Waters et al., in press). We have found that, when tested on the Daneman and Carpenter task with simple sentence structures that they have no difficulty understanding in isolation, all the DAT pa-

https://doi.org/10.1017/S0140525X99321786 Published online by Cambridge University Press

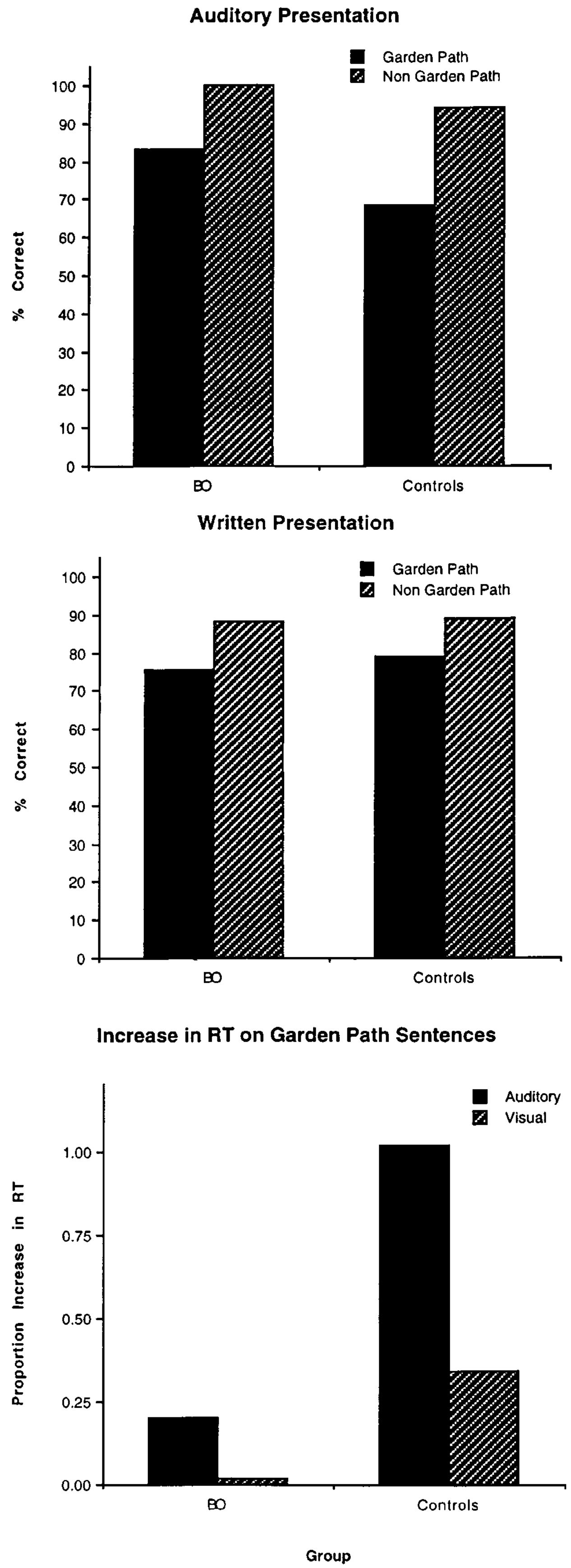


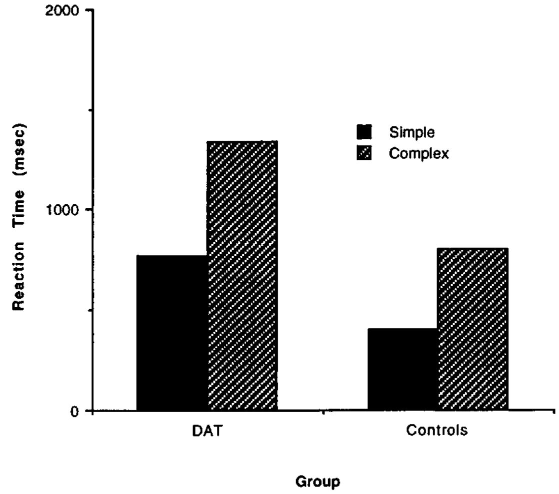
Figure 5. Performance of patient B. O., with a short-term memory impairment, and matched control subjects on acceptability judgment tasks with garden path sentences.

tients we have tested have had Daneman and Carpenter spans of one or less. Thus, the working memories of these patients are more reduced than those of normal subjects in whom the relationship between working memory and syntactic processing has been tested. Moreover, these patients' performances on working memory tests have been stable or have deteriorated over time (Baddeley et al. 1991). If the SR model is correct, then DAT patients should have particular difficulty with syntactically more complex sentences, and they should perform differentially poorly on the more difficult syntactic structures when a concurrent memory load is imposed. Our work has not found these results in this population.

In our first study, we tested 22 DAT patients and age- and education-matched controls for their ability to comprehend nine different syntactic structures using a sentence–picture-matching test (Rochon et al. 1994). The stimulus materials were similar to those used in the sentence–picture-matching and object-manipulation tasks in the dual-task experiments described above and were designed to contrast sentences that were matched for length and other relevant variables but that differed in syntactic complexity along with sentences that were matched for syntactic complexity but differed in terms of number of propositions. The results showed an effect of group, with DAT patients performing more poorly than controls, and a group-by-sentence-type interaction. Analysis of this interaction showed that DAT patients did not perform more poorly on the syntactically more complex sentences but, rather, that their performance was poorer than controls on sentences with two propositions. This pattern of an absence of an effect of syntactic complexity for DAT patients on sentence–picture matching has been replicated in several studies (Rochon & Saffran 1995; Waters et al. 1995; in press). We have also found the effect in speeded acceptability-judgment tasks. Figure 6 shows the results of one experiment from our laboratory comparing performance of DAT patients and matched controls on two-proposition simple and complex sentences in an acceptability-judgment task. In this task, both patients and controls showed an effect of syntactic complexity on reaction times. However, the effect was not greater in the patients than in the control subjects.

Patients with Parkinson's disease (PD) also have impairments in executive functions (Brown & Marsden 1988; 1991; Lees & Smith 1983; Taylor et al. 1986). We tested 17 nondemented PD patients and age- and education-matched elderly controls on a battery of tasks that tapped components of the STM system (articulatory rehearsal and phonological storage), verbal working memory capacity (reading span, self-ordered and externally ordered number generation), and various aspects of executive functions (Stroop color–word interference, Wisconsin card sorting, verbal fluency). Subjects were then tested for the comprehension of sentences differing in syntactic complexity and number of propositions, using a sentence–picture-matching task. Verbal working memory spans were significantly reduced in PD patients compared to controls, but rehearsal and storage functions were normal in span tasks in these patients. The performance of PD patients differed from that of controls on the sentence comprehension task, but only on sentences that contained more propositions (Fig. 7). Comparisons of sentences that differed in syntactic complexity but that held other factors constant, such as subject- versus object-relativized sentences, were not significant.

https://doi.org/10.1017/S0140525X99321786 Published online by Cambridge University Press



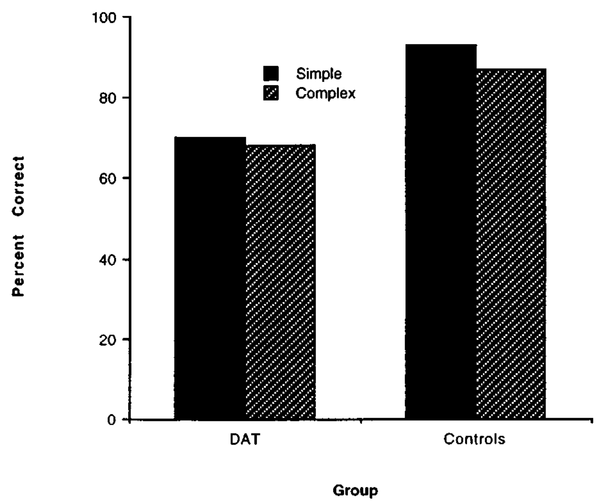


Figure 6. Performance of patients with dementia of the Alzheimer type (DAT) and matched normal controls on a sentence-acceptability judgment task. Simple, syntactically simple sentences; complex, syntactically complex sentences.

These data suggest that, as with DAT patients, PD patients do not have impairments in structuring sentences syntactically.

We have also found that the ability of DAT and PD patients to structure sentences syntactically is not affected any more than that of normal subjects by a concurrent memory load. In one study, DAT patients performed the sentence–picture-matching task outlined above in a no-interference condition and while retaining a concurrent memory load that was equivalent to their span or one less (Waters et al. 1995). Overall performance was poorer with the digit load, but comparisons of length-matched sentences showed that the concurrent memory load exacerbated the effect of number of propositions, but not the effect of syntactic complexity, in the patient group. We have repeated this study with the PD group described above. As with the DAT patients, the concurrent memory load did not exacerbate the effect of syntactic complexity in the PD group (Fig. 8).

These studies used untimed accuracy measures of sentence comprehension. Although they are not subject to the

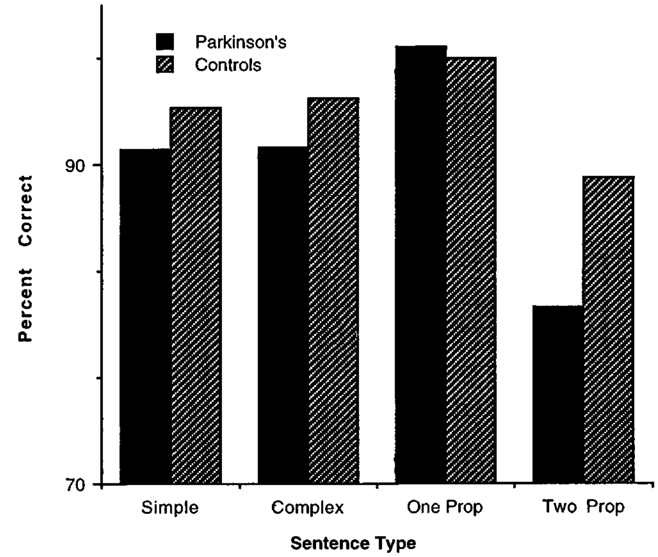


Figure 7. Performance of patients with Parkinson's disease (PD) and matched controls on a sentence–picture-matching task. Simple, syntactically simple sentences; complex, syntactically complex sentences; One Prop, sentences with one proposition; Two Prop, sentences with two propositions.

criticism that effects of syntactic structure were missed because of ceiling effects (because some comparisons of sentences that differed in syntactic complexity were made on pairs of sentences on which performance was below ceiling), measures of on-line processing may be more sensitive than an accuracy result. Using a cross-modal naming task to study on-line processing in DAT patients, MacDonald and her colleagues found that DAT patients were as sensitive as age-matched controls to grammatical violations and as able

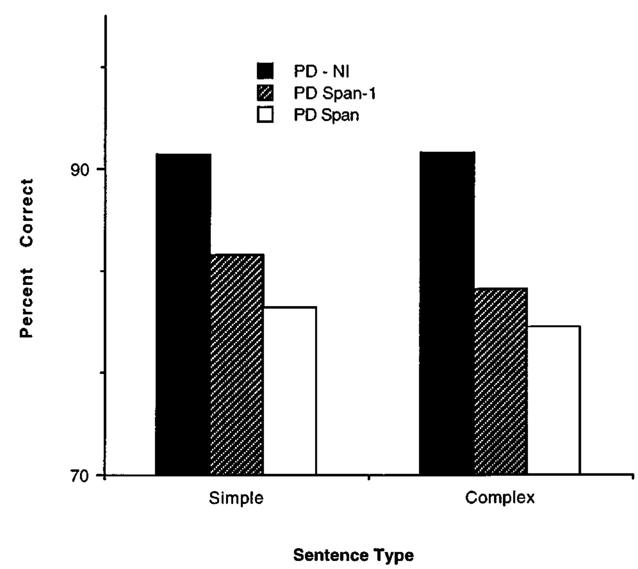


Figure 8. Performance of patients with Parkinson's disease (PD) on a sentence–picture-matching task with and without concurrent digit loads. Simple, syntactically simple sentences; complex, syntactically complex sentences; NI, no interference; Span-1, concurrent one-less-than-span condition; Span, concurrent span condition.

as age-matched controls to use frequency information and semantic and syntactic contexts to resolve syntactic ambiguities (Lalami et al. 1996; Stevens et al. 1996). The entire pattern of performances suggests that DAT patients have relatively well-preserved abilities to structure sentences syntactically in sentence interpretation. Preliminary results in our laboratory indicate that the same appears to be true of PD patients: A group of 12 PD patients showed the same pattern of performance as normal controls on the auditory moving-windows task.

We have combined the use of timed measures with dual-task conditions in a study of the ability of patients with DAT to make speeded acceptability judgments to auditorily presented sentences that varied in their syntactic complexity under no-interference and two concurrent tracking conditions (Waters & Caplan 1997). In a dot-tracking condition, subjects were required to press the switch on a mouse that corresponded to the location of a dot that appeared on the computer screen (left, middle, or right). In a digit-tracking condition, a continuous random sequence of the digits "1," "2," and "3" was presented in the center of the computer screen and subjects indicated which digit had been presented by pushing one of three switches on the mouse. DAT and control subjects' error rates were equated on each of the secondary tasks by varying the duration of the dot and digit stimuli and the interval between these stimuli. Patients showed an effect of syntactic complexity in the RT and sentence accuracy measures, which did not increase under the tracking conditions. Patients also showed an effect of condition, which was not greater for the more syntactically complex sentences. Both patients and controls showed an effect of syntactic structure in the secondary task-performance measures, which was not larger for the patients. Hence this experiment found that a concurrent task failed to affect processing of syntactically complex sentences more than syntactically simple sentences under very demanding concurrent task conditions in low-capacity subjects, even when sensitive timed measures of syntactic processing efficiency were employed. Our experiments with DAT and PD thus indicate that patients with severely reduced working memory capacity retain extremely good syntactic processing capacities in sentence comprehension. We note, however, that some authors have asserted that sentence comprehension is impaired in patients with DAT (Emery 1985; Kontiola et al.1990; Tomoeda et al. 1990) and PD (Grossman et al. 1991; 1992; Lieberman et al. 1990; Natsopoulos et al. 1991). Examination of these studies suggests that DAT patients might have performed poorly because of impairments they have with aspects of post-interpretive processing, such as deficiencies in their ability to access semantic knowledge, to enact responses, and to accomplish other post-interpretive requirements of many of these tasks (see Rochon et al. 1994, for discussion). For instance, real-world knowledge is necessary to understand and to respond to questions that are posed in some comprehension tasks (Emery 1985; 1986), and DAT patients might have done poorly in these tasks because of impairments affecting semantic memory even if they understood the stimulus sentences. Enactment is a response requirement of many of the tasks used (e.g., the token test; DeRenzi & Vignolo 1962), and it requires visuospatial, perceptual, and practice abilities that are often abnormal in DAT patients (Christensen 1974). Memory abilities are also required in many of the tasks used in existing studies (Hart 1988). The studies that show little or no sentence-comprehension impairment in DAT patients have tended to use tasks with simpler demands, such as sentence–picture matching (Rochon et al. 1994; Schwartz et al. 1979; Sherman et al. 1988; Smith 1989; but see Grober & Bang 1995). The same observations hold true regarding studies in which sentence-comprehension impairments have been found in PD patients (Grossman et al. 1991; 1992; Lieberman et al. 1990; Natsopoulos et al. 1991). For instance, Natsopoulos et al. (1991) had PD patients match a sentence to one of six pictures. Poor performance on these tasks could therefore reflect problems in handling these task demands.

We have suggested at several points in this target article that the effect of the number of propositions in sentence–picture matching, enactment, and acceptability-judgment tasks arises at the post-interpretive stage of sentence processing. Our finding of an increased effect of the number of propositions in DAT is consistent with the view advanced above that these patients' poor performance on some sentence comprehension tasks is due to their having impairments with such processes. Because of the importance of the claim that the effect of number of propositions reflects post-interpretive processing, we have examined the effect of manipulating the nonlinguistic visual and memory demands of the sentence–picture-matching task on the effect of syntactic complexity and number of propositions in DAT patients and controls (Waters et al., in press).

In one experiment, subjects matched a spoken sentence to one of two pictures that appeared either before or immediately following the presentation of the sentence. The second experiment used a video-verification task in which subjects were required to determine whether a spoken sentence matched a videotaped depiction of the action in the sentence or a syntactic foil. In this task, the spoken sentence sometimes ended before the action was completed, requiring the propositional content of the sentence to be maintained in memory while the action in the video unfolded. In the third experiment, in different conditions, subjects were required to determine whether a spoken sentence matched a single picture or to choose the picture that matched the sentence from an array of two or three pictures. In all tasks, DAT patients were affected by the number of propositions in the presented sentence, but not by the syntactic complexity of the sentence. Comparison across the one-, two, and three-picture versions of the task showed that the magnitude of the effect of number of propositions increased for both the DAT and control subjects as the number of pictures in the array increased (Fig. 9). In addition, analysis of the data from each of the tasks separately showed that the effect of number of propositions did not occur when the foil depicted an incorrect lexical item but only when subjects were attempting to match the target to a foil that required a syntactic analysis by depicting reversed thematic roles. These results support the view that the effect of number of propositions arises at a post-interpretive stage of processing at which the thematic roles in the sentence are held in memory and matched against an analysis of a picture.

Finally, in one additional study, we have found evidence that the effect of number of propositions seen on the sentence–picture matching task is related to the functioning of a general-purpose verbal working memory system. We examined the relationship between the magnitude of the proposition effect and DAT patients' performance on tasks that assessed primary memory (digits forward, Corsi

https://doi.org/10.1017/S0140525X99321786 Published online by Cambridge University Press

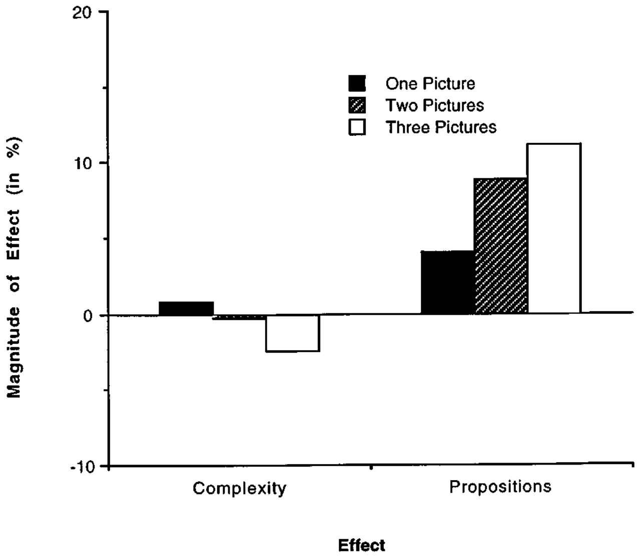


Figure 9. Average combined performance of patients with dementia of the Alzheimer type (DAT) and normal subjects on sentence–picture matching with one-, two-, and three-picture arrays, as a function of syntactic complexity and number of propositions.

block, auditory and visual word span), articulatory rehearsal/ phonological storage processes in STM (auditory and visual phonological similarity and word length), and measures of working memory function (digits backward, working memory span, dual-task performance). The magnitude of the effect of number of propositions was unrelated to all of the measures of primary memory and phonological storage/articulatory rehearsal (except visual word span) but was significantly correlated ($r = -.62$–$.70$) with all three working memory measures.

To summarize the results of our studies of the number-of-propositions effect, we have found that: (1) "intrinsic" limitations in working memory capacity in normal subjects (in whom the measurement of working memory capacity may be inaccurate) and in patients with DAT and PD are associated with an increase in the number-of-propositions effect in several tasks; (2) an "extrinsic" verbal memory load imposed by a concurrent task often increases the number-of-propositions effect; and (3) the magnitude of the number-of-propositions effect correlates negatively with measures of working memory in DAT patients. This set of findings provides support for the view that the number-of-propositions effect arises at a stage of processing that shares resources with working memory tests such as the reading span task. In addition, increases in nonlinguistic task demands such as the presence of a larger number of pictures in sentence-picture matching increases the number-of-propositions effect. This suggests that these resources are deployed at the post-interpretive stage of sentence processing.

We conclude this section with a brief note on sentence production in patients with DAT. Although not directly concerned with sentence comprehension, the results of several studies of this function are suggestively similar to those we have reviewed in the domain of comprehension.

Fluent, syntactically well-formed speech typically characterizes the conversational output of patients with DAT and has been found in DAT patients in picture description (Hier et al. 1985; Kemper et al. 1994) and sentence construction tasks (Schwartz et al. 1979), suggesting to many researchers that syntactic abilities are preserved in these patients, at least in the early stages of the illness (Bayles 1982; Blanken et al. 1987; Hier et al. 1985; Illes 1989; Irigaray 1973; Kempler et al. 1987; Kirshner 1982; Schwartz et al. 1979). DAT patients have also been found to correct errors of syntax and phonology, but not semantic errors, in anomalous sentences (Bayles 1982) and to make better use of syntactic than semantic cues in disambiguating spoken homophones while writing them to dictation (Kempler et al. 1987; Schwartz et al. 1979).

Although DAT patients have been shown to produce both oral and written language that is syntactically well structured, the language production of these patients does differ from that of age-matched controls in other important respects. In one study, more than 350 DAT patients were asked to write a single sentence and these productions were scored for their length in words and clauses as well as for the number of propositions produced (Kemper et al. 1993). The sentences were also scored for six categories of grammatic constituents, including pronouns, main verbs, secondary verbs, negatives, conjunctions, and questions. Results showed that 89% of the variance in the clinical rating of dementia severity could be accounted for by sentence length in clauses and propositional content. This suggests that the ability to produce propositions in written form is reduced as the severity of dementia increases. Furthermore, estimates of these linguistic abilities early in life appear to be powerful predictors of cognitive function and Alzheimer's disease in later life. In a longitudinal study, Snowdon et al. (1996) analyzed autobiographies written by 93 nuns when they entered a convent and evaluated the cognitive performances of these women when they were between 75 and 95 years of age (a span of some 58 years). Alzheimer's disease was assessed neuropathologically in a sample of 25 participants who died. A stronger and more consistent association was found between cognitive function later in life and the density of propositions in these early narratives than between cognitive functions and the grammatical complexity of the narratives. Strikingly, low idea density was present in the autobiographies of 90% of women with neuropathologically proven Alzheimer's disease but in only 13% of those without Alzheimer's disease. Further studies have shown that measures of idea density and grammatical complexity are highly stable over the life span ($r = .62$–$.74$) (Kemper et al. 1996). The authors hypothesize that low linguistic ability early in life may be an early expression of Alzheimer's disease neuropathology. From our point of view, these results suggest a finer distinction. The greater relationship between measures of propositional density and DAT than between measures of grammatical well formedness and DAT is consistent with the view that these patients are impaired in formulating concepts but are able to use the forms of language to convey the concepts that they do activate. This is a division of function related to language production that corresponds to that we have suggested between post-interpretive and interpretive aspects of the comprehension process.

3.3. Studies of patients with reduced resources for syntactic processing. We may also approach the question of the relationship between working memory capacity and syntactic processing in sentence comprehension from an-

https://doi.org/10.1017/S0140525X99321786 Published online by Cambridge University Press

other angle, by examining the effect of a concurrent memory load on syntactic comprehension in patients in whom there is evidence for a reduction in the resources available for syntactic processing in sentence comprehension. Patients with aphasia provide such cases.

Research into the nature of the sentence comprehension impairments seen in patients who are aphasic subsequent to a left hemisphere stroke has shown that many such patients have disturbances affecting their ability to use syntactic form to determine the meaning of a sentence (Caplan et al. 1985; 1996; see Berndt et al., 1996, for review). Several aspects of the performance of these patients suggests that one reason for this impairment is a reduction in the processing resources that a patient can apply to this task. One piece of evidence that favors this view is that groups of patients have been shown to have difficulty understanding sentences with more complex syntactic structures (Caplan et al. 1985). Second, factor analyses have shown that a single factor on which all sentence types load accounts for about two-thirds of the variance in many syntactic-comprehension tasks in aphasic groups (Caplan et al. 1985; 1996; 1997). Third, cluster analysis shows that patients tend to be grouped according to their overall level of performance in these tasks, with performance in more impaired clusters showing greater effects of syntactic complexity (Caplan et al. 1985; 1997). Finally, some patients have been able to interpret sentences when either of two syntactic features was present, but not when both were found in a sentence (Hildebrandt et al. 1987). These patterns of performance are consistent with the hypothesis that the problem in syntactic processing in sentence comprehension seen in many aphasic patients results in part from reductions in their ability to allocate processing resources to the syntactic-comprehension task.

According to SR theory, any reduction in resources available for syntactic processing comes from a pool that is shared with other verbal tasks and, therefore, syntactic complexity effects will be increased in aphasic patients under a concurrent verbal memory load condition. In SSIR theory, these pools are separate and there is the strong prediction that aphasic patients will not show an increase in syntactic complexity effects under a concurrent load condition.

We examined the effect of a concurrent digit load on the sentence-comprehension performance of aphasic patients (Caplan & Waters 1996). More than 200 aphasic patients were screened to ensure that we tested only patients who showed effects of syntactic complexity on a sentence–picture-matching task, whose performance was below ceiling and above chance on that task, and whose abilities to repeat single words permitted them to be tested on a digit-span task. We selected ten patients who met these criteria, and tested them on a sentence–picture-matching task in no-interference and concurrent load conditions (span and span minus one). Although aphasic patients showed large effects of syntactic complexity when tested on the sentence–picture matching test without a concurrent load, these effects were not exacerbated by the addition of a memory load. Their performance on the concurrent memory task was poorer with larger digit loads, but the effect of syntactic complexity was not exacerbated. In the digit-recall data, there was an effect of number of propositions but not of syntactic complexity. These results provide striking evidence for the separation of the resources used in syntactic processing in sentence comprehension and those required for span tasks.

3.4. Summary of studies with patient populations. The studies reviewed in this section provide important evidence against the single processing resource model. Patients with several etiologies of CNS disease who have reduced verbal working memory have been shown to do well on tests of syntactic comprehension. Where effects of syntactic structure arise in these groups, the effects are comparable to those found in normal subjects. Concurrent verbal loads do not disproportionally affect comprehension of syntactically more complex sentences in these patient groups. These results are based on patients' performances on demanding concurrent tasks and include measurements of RTs as well as accuracy. They have documented a domain of retained functional capacity in the midst of the many cognitive and executive limitations found in these patients. In the aphasic population, for whom several arguments can be made that there is a reduction in processing resources used for syntactic comprehension, a concurrent verbal memory load does not exacerbate the effect of syntactic complexity. All these results are consistent with the view that the resources used in syntactic processing in sentence comprehension are not reduced in patients with reduced verbal working memory capacity and are not shared by the digit-span task. In contrast, patients with reduced working memory capacities have shown larger effects of the number of propositions in a sentence than control groups, and these effects were sometimes increased under concurrent-load conditions. This suggests that the ability to match a proposition to a picture or to check its plausibility against information in semantic memory is affected in these patients, and is affected by a concurrent digit-span task. Although other accounts must be considered, this pattern of results is consistent with the view that the reduction in working memory seen in these patients affects these post-interpretive processes and that the resources used in these post-interpretive tasks are shared by immediate serial-recall tasks.

4. Discussion: Fractionating verbal working memory

The studies reviewed above provide strong evidence that subjects whose verbal working memory capacity is reduced on standard tests of this function can retain the ability to use syntactic structure to determine sentence meaning. This is true not only in normal subjects (i.e., those without neurological disease), in whom the measurement of working memory may be unstable over time and lead to misclassification, but also in patients with Alzheimer's disease and Parkinson's disease, whose verbal working memory and executive control functions are significantly impaired compared to those of normal subjects. In many experimental paradigms, the processing of syntactically more complex sentences is not disproportionately affected by a concurrent verbal memory load, either in normal subjects or in subjects with extremely reduced working memory capacity or in aphasic patients. These results provide evidence that the working memory system involved in sentence interpretation is separate from that measured by standard tests of working memory.

If there is a specialization within verbal working memory for the assignment of syntactic structure and its use in sentence interpretation, this specialization may not be restricted to this one aspect of the sentence interpretation process. Syntactic processing is one of a set of related operations that transform the acoustic signal into a discourse-coherent semantic representation in normal, everyday conversations. A partial list of other types of operations in this process includes acoustic–phonetic conversion, lexical access, recognition of intonational contours, and determination of discourse-level semantic values such as topic, focus, coreference, causality, and temporal order of events. These operations normally act in concert: a listener computes lexical, propositional, and discourse meanings in every normal communicative act (Marslen-Wilson 1987). Most researchers consider that the integrated set of operations that activate representations in sentence and discourse interpretation has processing characteristics that are part of its usual successful functioning. Processors are thought to be obligatorily activated when their inputs are attended to. They generally operate unconsciously, and they usually operate quickly and remarkably accurately. The operations of the language interpretation process are thus integrated in two senses: they always compute items within the same restricted set of representational types, and they do so in a particular manner. Moreover, they are among the most highly practiced of human cognitive functions. Because of their integration and degree of overpractice, we have suggested that one resource system is utilized by all these different types of processes that combine in the interpretation process (Caplan & Waters 1996; Waters & Caplan 1996a). Because of its greater inclusiveness, we will refer to this hypothesis as the "separate *language* interpretation resource" (SLIR) hypothesis, an extension of the "separate *sentence* interpretation resource" (SSIR) hypothesis.

According to the SLIR hypothesis, the entire set of operations that compute a coherent discourse meaning depends on a single working memory resource. It would therefore be supported by the finding that the efficiency of these operations is positively correlated across normal subjects and that the effects of increasing the working memory demands of different interpretive operations are interactive. Operationally, the entire system need not be tested for this hypothesis to begin to be investigated; these predictions can be applied to pairs of operations. For example, the theory could begin to be tested by investigating the effect of variables that affect lexical access, such as frequency, and of variables that affect sentence interpretation, such as syntactic complexity. In the model, the efficiency of post-interpretive operations is not related to that of interpretive operations. It therefore predicts that the efficiency of interpretive and post-interpretive operations is not correlated and that the effects of increasing the working memory demands of interpretive and noninterpretive operations are not interactive. The SLIR hypothesis predicts that working memory capacity, as measured on a task that emphasizes controlled, conscious manipulation of verbal information, will not correlate with processing efficiency for any component of the interpretation process, that a concurrent verbal memory load will not interfere with more demanding processing within any interpretive component, and that each language-processing component involved in interpretive processing will be preserved in some neurological subjects with reduced working memory.

The view that we have presented for an integrated set of language processing operations in the interpretation process has been challenged (Just et al. 1996a). One argument that has been made is that the comprehension process is not always as automatic and seamless as we have represented it to be. The sentence with which we began this paper, *The man that the woman that the child hugged kissed laughed,* can be interpreted, but its successful interpretation is not usually obligatory, unconscious, and fast. Reanalysis of ambiguous lexical items (*We hated the cheap hotel room because of all the bugs we saw in it. We realized our conversations would not be private)* and of syntactic structures (*The aggressive trial lawyer questioned in minute detail by the judge hesitated)* also often involves conscious, slow, controlled processing. So does drawing certain types of complex inferences (*Harvey frequently invested in junk bonds. He now sells pencils at the corner of Broad and Main*), or revising inferences (*John returned home from the auction $500 poorer. He swore he would never attend a function in a rough part of town again*). Listeners appear to accomplish sentence interpretation in two ways, in the usual obligatory, unconscious and fast mode, or in a mode in which conscious, controlled processes are applied to the task. We do not conceive of the latter type of processing as belonging to the set of operations that we suggest utilize a specialized resource pool. There have been several models of the change from one processing mode to another, especially in the area of syntactic reanalysis (see, e.g., Sturt & Crocker 1996). A research program could be directed toward the question of whether the divisions made by these models predict the nature of the working memory system involved in processing different types of sentences.

A second issue that has been raised is that the interpretation process – specifically, syntactic processing – is influenced by nonlinguistic factors such as pragmatic expectations (Trueswell et al. 1994) and the frequency with which particular lexical items co-occur in a language (MacDonald et al. 1994; see papers in MacDonald 1997). These findings do not undermine the analysis presented here. The fact that the interpretation process can accept such information as input does not entail that it cannot be distinguished from other verbally mediated functions. One way to distinguish it is on the basis of the *combination* of its input, its intermediate representations, and its output compared to those of post-interpretive tasks. We may think of interpretive processes as a function that maps the acoustic signal onto a representation of the preferred literal meaning of sentences in a coherent discourse, that computes an intermediate set of linguistic representations (phonemes, words, syntactic structures, etc.). No other function accomplishes this mapping. Most post-interpretive processes are functions between the output of this process and logical entailments, confirmations of the presence of a proposition in a memory system, plans for action, or some other endpoint. Very few functions other than language interpretation compute any of the intermediate representations that are computed by the interpretation process, and none computes all the representations that are routinely computed by this process. In short, encapsulation in Fodor's (1983) terms was only one hypothesized feature of the interpretation process. The interpretation process does not have to be encapsulated to be cognitively separable or to rely on a specialized resource pool (Caplan 1985).

If we tentatively accept the view that there is a specialization in the verbal working memory for language interpretation, we may try to place this specialization within a neurobiological context. One can only speculate about the origin of such a specialization. It may be the product of innately determined neural development (Rakic 1988), practice (Grafton et al. 1992; 1995), or both. No data are presently available that determine this issue.

Somewhat more is known about the neural systems that may support such a specialization, although knowledge in this area is also very incomplete. There is strong evidence that the dominant perisylvian association cortex is critical for all language-interpretive functions (Caplan 1987). This area is the best candidate for the location of the neural system that supports the processing resources used in language interpretation. As regards specifically syntactic processing in sentence comprehension, lesions throughout this region affect syntactic processing in sentence comprehension in ways that suggest a reduction in available processing resources (Caplan et al. 1985; 1996; 1997). It has been suggested that one region in the left perisylvian cortex – the pars opercularis and traingularis of the third frontal convolution, Brodmann's areas 44 and 45 (Broca's area) – might be particularly important in this function. Evidence supporting this localization comes from the finding that on-line processing of sentences with relative clauses is abnormal in patients with Broca's aphasia, who tend to have lesions that involve this region, but not in patients with fluent aphasias, whose lesions tend to spare this area (Swinney & Zurif 1995; Swinney et al. 1996; Zurif et al. 1993). Studies of event-related potentials have also identified an early negative wave in the left frontal region associated with processing object-relative clauses (Kluender & Kutas 1993; Neville et al. 1991), and two studies have shown a localized increase in regional cerebral blood flow (rCBF) in the pars opercularis when subjects made acceptability judgments about sentences with object- as opposed to subject-relative clauses (Caplan et al., in press; Stromswold et al. 1996). These results are consistent with a narrower degree of localization of processing resources used in at least one aspect of syntactic comprehension in Broca's area. Not all available data support this localization, however. Just et al. (1996b) have reported increased rCBF in both the left frontal and the left temporal lobes (and, to a lesser degree, in the homologous contralateral cortical regions) in a question-answering task using sentence types that were similar to those tested using acceptability judgment by Stromswold et al. (1996). These differences across studies have not yet been explained. Overall, the picture that emerges is that association cortex in the dominant perisylvian region is supported in all likelihood by sustaining projections from thalamus and basal ganglia and probably also the contralateral nondominant perisylvian association cortex, is responsible for language interpretation, and is the best candidate for the locus of the resources that are used in this aspect of language processing.

In contrast, many studies show that working memory tasks that involve conscious, controlled processing of language representations activate dorsolateral frontal regions of the left hemisphere rostral to the perisylvian association cortex (for representative results, see Jonides et al., 1993, and Petrides et al. 1993). The division of neural structures into a perisylvian cortical region involved in language interpretation and a more anterior region that supports conscious, controlled verbal working memory functions is consistent with the division within the verbal working memory system that we have advocated here.

It is worth noting that there are areas of the brain that may be involved in both types of tasks. For example, the anterior cingulate gyrus is activated in a wide variety of tasks – both those that involve language interpretation and those that involve conscious, controlled verbal working memory – and has been hypothesized to be a structure involved in setting subjects' level of arousal, thereby allowing them to devote resources to a task (Posner et al. 1988). SLIR theory does not deny the existence of such regions, but its support comes from the existence of regions that contribute to working memory in only one of these types of language functions.

In summary, we have presented evidence that the use of processing resources in assigning syntactic structure and using that structure to determine the meaning of a sentence is separable from a subject's verbal working memory capacity as measured by standard working memory tasks. We have accordingly argued that the working memory system contains specializations for different verbal processes and that one such specialization is used for the integrated processes involved in the determination of the meaning of discourse. Our hypothesis requires additional specification that can come only as models of both language processing and working memory become better developed and justified, but it can be tested in the context of present-day knowledge about both language processing and working memory.

ACKNOWLEDGMENTS

This research was supported by grants from the National Institute of Aging (AG09661) and from the Medical Research Council of Canada (MA9671) to David Caplan and Gloria Waters.

NOTES

1. These predictions depend on the assumption that the slope of the relationship between resource demands and sentence complexity is monotonically increasing. This simplifying assumption has been made throughout the literature on sentence processing and working memory. We make it in interpreting the results presented below, recognizing that it will have to be empirically validated or that more complex models will have to be adopted and tested.

2. This conclusion depends on the size of the resource pools in a separate-sentence-interpretation-resource model being uncorrelated. However, although the possibility of correlated sizes of resource pools provides an "out" for the separate-sentence-interpretation-resource model in the case that there are interactions between sentence complexity and working memory capacity group, it is fair to say that such interactions still favor the single-resource model, at least by making the separate-resource model adopt more complex assumptions.

3. Span is often contrasted with concurrent articulation (repeating a single word such as "the" or "double"), which is thought only to interfere with rehearsal, not to compete with another task for central executive resources (Baddeley & Hitch 1974; Just & Carpenter 1992).

4. These predictions also depend on assumptions about floor and ceiling effects in experimental results. For instance, if the combination of task and sentence-processing demands is less than the working memory resources available to even low-span subjects, ceiling effects would be expected to eliminate interactions between load, sentence type, and span group. Much more complex patterns of results can be predicted if other assumptions about floor and ceiling effects are made (see Waters & Caplan 1996a, footnote 3).

Open Peer Commentary

Commentary submitted by the qualified professional readership of this journal will be considered for publication in a later issue as Continuing Commentary on this article. Integrative overviews and syntheses are especially encouraged.

Complexity effects are found in all relative-clause sentence forms

Glenda Andrews[a] and Graeme S. Halford[b]

[a]*School of Humanities and Social Sciences, Charles Sturt University, Wagga Wagga, NSW, Australia 2678;* [b]*School of Psychology, University of Queensland, Brisbane, Queensland Australia 4072.*
gandrews@csu.edu.au **gsh@psy.edu.au**
www.psy.uq.edu.au/people/department/gsh/index.html

Abstract: We argue that if a different definition of sentence complexity is adopted and processing capacity is assessed in a way that is consistent with that definition, then the Caplan & Waters distinction between interpretive versus postinterpretive processing is unnecessary insofar that it applies to the thematic role assignment in relative-clause sentences.

Caplan & Waters (C&W) propose a distinction between interpretive and postinterpretive processing. The distinction implies that determining *who did what to whom* by assigning nouns to their thematic roles imposes little or no processing load in relative clause sentences such as sentences (1) and (2), but does impose a load in sentences such as sentence (3) (sect. 4).

1. The duck that the monkey touched walked.
2. The artist that the waiter warned the chef about talked.
3. The clown that the teacher that the actor liked watched laughed.

Our claim is that loads are imposed by all three sentence types but that their magnitudes differ. We quantify processing loads in terms of Halford et al.'s (in press) domain general metric of relational complexity. "Relational complexity" refers to the arity of relations, that is, the number of arguments or entities that are related. Each argument corresponds to a dimension and an N-ary relation is a set of points in N-dimensional space. The number of dimensions corresponds to the number of interacting variables that constrain responses. Unary relations have a single argument as in class membership, dog(Fido). Binary relations have two arguments as in larger-than(elephant, mouse). Ternary relations have three arguments as in addition(2,3,5). Quaternary relations such as proportion have four interacting components as in 2/3 = 6/9. Quinary relations entail five interacting components. Thus a metric of relational complexity is defined, and processing load is predicted to increase with complexity. Tasks that require more than a quaternary relation to be processed in parallel are predicted to impose loads that exceed the capacity of the average adult. When capacity limits are exceeded, individuals will attempt to use segmentation and chunking strategies to reduce complexity.

We have used this metric to quantify the complexity of relative-clause sentences. Complexity corresponds to the number of role assignments that must be made in the same decision. Because of the nonstandard order in which the roles occur in center-embedded (CE), object-relative sentences and their low frequency of occurrence, most people lack strategies for segmenting the sentences so that role assignments can be made sequentially. Consequently, the assignments are likely to be made in the same decision. Sentences (1), (2), and (3) require, respectively, three, four, and five role assignments and have the complexity of ternary, quaternary, and quinary relations. Right branching (RB), subject-relative sentences such as sentences (4), (5), and (6), which occur more frequently and are easily segmented because the roles occur in standard order, serve as controls for number of role assignments and semantic content.

4. The monkey touched the duck that walked.
5. The waiter warned the chef about the artist that talked.
6. The actor liked the teacher that watched the clown that laughed.

We used visual, self-paced presentation of sentences with probe questions such as Who *verbed?* Who was *verbed?* What did the *noun* do? to assess comprehension in normal adult participants. Results were consistent with the complexity metric in that comprehension difficulty increased with number of roles, but this effect was greater for CE than for RB sentences.

We focus here on the capacity–comprehension relationship. We included a reading span task reflecting Just and Carpenter's (1992) working memory approach and an n-term premise-integration task based on Halford et al.'s (1998) relational complexity approach. The n-term test requires reasoning about ordinal relations between elements. The items varied in complexity as defined above. We found a significant correlation of reading span with four-role CE sentences, $r = .41$, $p < .05$, but not with three-role CE or five-role CE sentences, $r = .15$ and $r = .11$, respectively. N-Term performance was significantly correlated with three-role CE, $r = .24$, $p < .05$ (one-tailed), four-role CE, $r = .45$, $p < .001$, and five-role CE sentences, $r = .48$, $p < .001$. These patterns are not consistent with C&W's distinction, which would predict significant capacity–comprehension correlations for five-role CE sentences only, because these require post-interpretive processing.

The n-term–comprehension correlations are consistent with the interpretation that understanding CE sentences entails processing of complex relational information. There was also evidence of a capacity × complexity interaction, which C&W claim is not generally found. The required interaction corresponds to a significantly higher correlation between capacity and comprehension for sentences of higher complexity than for sentences of lower complexity. Stiger's (1980) T_2 statistic showed that the correlation of n-term with the average of four- and five-role CE sentences ($r = .59$) was significantly higher than with three-role CE sentences (r = .24), $T_2(62) = 2.81$, $p < .05$. The correlation of reading span with four-role CE sentences was significantly higher than with three-role CE sentences, $T_2(62) = 1.98$, $p < .05$ (one-tailed). Note that this latter interaction does not involve five-role CE sentences, which C&W claim are the ones that require verbal working memory.

These interaction effects entail the manipulation of complexity in terms of number of role assignments made in a single decision. In the studies reviewed by C&W, the complexity manipulations typically involved comparing sentences with the same number of role assignments but that differ in form (e.g., CE object-relative vs. CE or RB subject-relative, see sects. 2.1 and 2.2). In our work, the capacity measures were sensitive to the form manipulation only if sentences required four or five roles assignments. The correlation of reading span with four-role CE sentences ($r = .41$) was significantly higher than with four-role RB sentences ($r = .14$), $T_2(62) = 2.08$, $p < .05$. The correlation of n-term with five-role CE sentences was significantly greater than with five-role RB sentences ($r = .15$), $T_2(62) = 2.08$, $p < .05$. Observation of significant capacity × complexity interactions seems to require consideration of number of role assignments in addition to sentence form. In defining complexity in terms of the number of roles assigned in the same decision, the relational complexity approach achieves this in a way that incorporates all relative-clause sentences.

The need to distinguish between relative clause sentences in which role assignment requires interpretive versus post-interpretive processing appears to be eliminated. However, our data do not preclude the usefulness of the distinction in other contexts.

https://doi.org/10.1017/S0140525X99321786 Published online by Cambridge University Press

Interfaces in memory

Zoltán Bánréti
Research Institute for Linguistics, Hungarian Academy of Sciences, Budapest, H-1250 Hungary. **banreti@nytud.hu**
www.nytud.hv/tlp/staff/banreti

Abstract: A distinction between interpretive processing and post-interpretive processing calls for a consideration of interface relations in systems of verbal memory. Syntactic movement of a phrase and the cognitive system of thought/mind interact. Systems of declarative memory and procedural memory interact.

Caplan & Waters (C&W) develop an interesting idea about the modular character of verbal working memory, but their central claim – that verbal working memory is composed of subsystems specialized for different types of verbal tasks – suffers from problems.

Displacement of a phrase and the cognitive system of thought/mind interact. Experimental results indicate that an English object-relativized sentence is more demanding for interpretive processing in verbal memory than a subject-relativized sentence. According to C&W, the same is not true for post-interpretive processing in verbal memory: there is no correlation between the efficiency of interpretive and post-interpretive processes (as in sentences 8 and 11). But an object relativized sentence can be more demanding, because it contains two displacements: a movement of the object phrase and a movement of the subject phrase. This "displacement property" is an important aspect of language; phrases are interpreted as if they were in a different position in the sentence, where similar items sometimes do appear and are interpreted in terms of natural local relations (Chomsky 1997). In generative grammar, movement rules express the displacement property. Why should language have the displacement property? According to Chomsky (1997), it is motivated by interpretive requirements that are externally imposed by our systems of thought/mind. Displacements can be described in terms of interpretation: new versus old information, topic/comment, specificity, foreground versus background information, the agentive force that we find in displaced position, and so on. "These seem to require particular positions in temporal linear order, typically at the edge of some construction" (Chomsky 1997, p. 17). Object-relativized sentences and subject-relativized sentences contain a displaced object or subject at the edge of structure, respectively. Differences in their structural complexity reflect differences in their interpretive features as imposed by the system of thought/mind. Hence the areas for a specialization in the verbal working memory for language interpretation are far from clear. A distinction between interpretive processing and post-interpretive processing calls for a consideration of interface relations.

Degree of modularity? C&W refer to data on processing garden-path sentences like (8)–(10). These data indicate that garden-path sentences are harder to process for both high- and low-capacity subjects, regardless of the animacy of the first noun. C&W claim that this provides evidence for similar degrees of modularity of syntactic processing in both high- and low-capacity subjects. We do not know why. The data referred to indicate only that garden-path sentences are harder to process than normal ones. These results can be expected in single resource (SR) and separate sentence interpretation resource (SLIR) theories as well. The data say nothing about degrees of modularity.

Systems of memory interact. C&W claim that experiments with patients who suffer from serious cognitive disorders such as dementia of the Alzheimer type (DAT) or Parkinson's disease (PD) also support the modularity of verbal working memory system. It is not clear what kind of relations there are between verbal working memory and declarative memory and procedural memory. Ullman et al. (1997) present evidence that the lexicon is part of declarative memory system and that grammatical rules are processed by a procedural system. They describe Alzheimer disease (AD) subjects with severe impairments in learning new facts and words and remembering old ones, with relative sparing of the learning and processing of motor and cognitive skills. They found AD patients to be relatively unimpaired in the processing of rule-governed word forms and in processing the syntax of sentences. According to Ullman et al. (1997), PD patients exhibit impairments in the learning of motor and cognitive skills. PD patients also have difficulty with grammar, especially with syntax. They have trouble understanding sentences, and their spontaneous speech is syntactically simple. Recognizing words and facts remains relatively unimpaired. PD and AD groups show complementary deficits in rules and lexicon. C&W claim that AD patients with impaired working memory retain good syntactic processing capacities. It is surprising that C&W fail to find substantive differences in performance in PD patients. How can PD patients with impaired grammar (syntax) preserve good syntactic processing capacities? How can *both* AD and PD patients be impaired in post-interpretive processing alike? Does it mean that specialization in working memory is independent of declarative memory and procedural memory? Surely impairments in procedural memory and/or declarative memory affect a subsystem in working memory.

Ullman et al. (1997) report no essential differences for declarative memory in storing either facts or words and no essential differences for procedural memory in the processing of motor/perceptual skills and processing the syntax of sentences. This correlation weakens the idea of linguistic specialization in working memory. The findings of Ullmann et al. and C&W are incompatible.

I think linguistic projection from declarative to procedural memory is possible and necessary. Lexical items are retrieved from declarative memory, but they have to be combined to form an utterance, so they must be "transferred" to procedural memory. Grammatical (agreement) features of these items in an utterance must be checked, and likewise in procedural memory. For syntactolexical integration as such, attaching the content words to the syntactic frame, interaction between the two memory systems is again necessary. This fits well with a view on working memory as involving the circuitry of frontal and parietal/temporal areas.

Not so fast: Domain-general factors can account for selective deficits in grammatical processing

Elizabeth Bates,[a] Frederic Dick,[a] and Beverly Wulfeck[b]
[a]*Center for Research in Language, University of California, San Diego, La Jolla, CA 92093-0526 and* [b]*Department of Communicative Disorders, San Diego State University, San Diego, CA 92182-1518.* **bates@crl.ucsd.edu crl.ucsd.edu/~bates/ fdick@crl.ucsd.edu wulfeck@crl.ucsd.edu www-rohan.sdsu.edu/dept/chhs/cd/wulfeck.html**

Abstract: Normals display selective deficits in morphology and syntax under adverse processing conditions. Digit loads do not impair processing of passives and object relatives but do impair processing of grammatical morphemes. Perceptual degradation and temporal compression selectively impair several aspects of grammar, including passives and object relatives. Hence we replicate Caplan & Waters's specific findings but reach opposite conclusions, based on wider evidence.

Passives and object relatives are more difficult for agrammatic aphasics to process and interpret than actives and object relatives. Caplan & Waters (C&W) ascribe this well-known fact to deficits in a domain-specific grammatical processor, with a syntax-specific pool of working memory resources. They base this conclusion on evidence involving comprehension of a small subset of English syntactic structures, in patients with agrammatic aphasia, patients without receptive agrammatism despite working memory deficits, and young adults stratified by memory span and/or tested under a digit load to simulate deficits in domain-general working memory. Against Miyake et al. (1994), who simulated receptive agrammatism in normal subjects with speeded visual stimuli, C&W do not find specific effects on "hard" sentence types as a function of digit load or memory span. We have replicated C&W's results, but we have also replicated those of Miyake et al., and we believe that C&W have moved too fast in reaching their broad conclusions about the autonomy of grammatical processing. Using a broader range of structures, languages, and patient groups, with controls tested under a broader range of adverse processing conditions, we conclude that *specific deficits in grammar can be explained without recourse to a domain-specific resource or processing device.* Our interpretation differs from that of Miyake et al., but it is similar in spirit.

First, the same hierarchy of difficulty (actives, subject relatives > passives, object relatives) has been observed in several languages and in many different populations, including Broca's aphasics, Wernicke's aphasics, anomics without expressive agrammatism, and individuals in the early stages of first- or second-language acquisition. The pattern is not unique to any form of aphasia or to any lesion site (Dick et al. 1998; Naucler et al. 1998).

Second, other facets of receptive agrammatism (deficits in the use of function words and grammatical inflections) have been observed in a broad range of patient populations as well as in normals subjected to a broad range of stressors. Published and unpublished studies from our laboratory have simulated selective deficits in morphology (with relative sparing of word order) in college students processing under a digit load, a partial noise mask, low-pass filtering, and/or auditory compression. These results hold true, in varying degrees (depending on the strength of each information type under normal conditions), in English, Italian, and German (Bates et al. 1994; Blackwell & Bates 1995; Kilborn 1991).

Third (and most relevant to C&W's claim), selective deficits in the processing of passives and object clefts have been demonstrated in English college students, but under conditions different from those adopted by C&W. Because they failed to demonstrate effects of digit load or working-memory capacity on the above sentence hierarchy, C&W conclude that syntactic processing is affected only by deficits within a syntax-specific pool of processing resources and not by reductions in working memory outside this domain (as claimed by Miyake et al., based on results with speeded presentation). We have shown that C&W and Miyake et al. are both right: college students tested under a digit load (a task that disrupts computation of subject–verb agreement and other inflectional phenomena in our laboratory) are unimpaired in their ability to process passives and object clefts (replicating C&W), but students tested under perceptual degradation and/or temporal compression are selectively impaired on precisely those sentence types (replicating Miyake et al. in the auditory modality). Furthermore, students tested with *both* compression and perceptual filtering produced superadditive results, greater than those we would expect from adding separate effects of compression and degradation alone, and identical to results for aphasic patients in the same paradigm (similar error rates and similar patterns of individual variation in a cluster analysis) (Dick et al. 1998).

We conclude that the specific challenges posed by passive and object relatives are not unique to a single aphasia type and can be explained without recourse to syntax-specific mechanisms or to damage involving specific lesion sites. We propose a domain-general account of the specific difficulties posed by low-frequency syntactic structures that differs from the working-memory proposal of Miyake et al., reflecting the effects of structural frequency on *encoding* (activation of stimuli) rather than *memory* (maintenance of stimuli in working memory). Grammatical morphemes are vulnerable to stressors of either type (including digits); low-frequency word orders are vulnerable at encoding but form solid memory traces that are mnemonically robust if they make it over the encoding threshold. This would explain why patients with working memory deficits do not show the predicted pattern, but it does not permit C&W to leap to a much stronger conclusion, namely, that syntactic deficits reflect damage to an autonomous processor, independent from the processing resources used by other cognitive systems. Our account makes differential predictions for the fate of complex sentence types under stress in cross-linguistic comparisons, results that are supported by preliminary findings for German and Italian.

ACKNOWLEDGMENTS
This work was supported by National Institute on Deafness and other Communication Disorders grant 2-R01-DC00216 and National Institute of Neurological Disorders and Stroke grant NS22343.

Fractionated working memory: Even in pebbles, it's still a soup stone

Morten H. Christiansen[a] and Maryellen C. MacDonald[b]
[a]*Department of Psychology, Southern Illinois University, Carbondale, IL 62901-6502 and* [b]*Program in Neural, Informational and Behavioral Sciences, University of Southern California, Los Angeles, CA 90089-2520.*
morten@siu.edu; mcm@gizmo.usc.edu
www.siu.edu/departments/cola/psycholfaculty/mhc.html
siva.usc.edu/~mcm/

Abstract: We agree with Caplan & Waters that there are problems with the single-resource theory of sentence comprehension. However, we challenge their dual-resource alternative on theoretical and empirical grounds and point to a more coherent solution that abandons the notion of working memory resources.

Caplan & Waters (C&W) argue for the inadequacy of the single-resource theory of verbal working memory in sentence comprehension. We are sympathetic to this position and see two approaches to developing an improved account. One is the approach that C&W adopt: dividing the single working memory resource into two independent resources, each dedicated to a particular processing module. The other is to reject the construct of a limited processing resource; Navon (1984) clearly articulated this position and likened the notion of processing resources to a theoretical "soup stone," contributing no explanatory power to theories of cognition. We have elsewhere advocated this position for theories of language comprehension and have argued that the abandonment of the notion of verbal working memory provides a superior account of the individual differences in sentence-comprehension abilities that C&W review (MacDonald & Christiansen 1998). We see C&W's dual-resource account as nothing more than cracking the superfluous soup stone into pebbles. Instead, we suggest that individual differences emerge from interactions between variations in linguistic experience (e.g., some people read more than others) and processing architecture (e.g., some people have more accurate phonological representations than others). Thus, individual differences in working memory tasks correlate with language comprehension not because there is a separate resource constraining these tasks but rather because the same architectural and experiential constraints that shape the accuracy of language comprehension also affect skill in performing verbal working memory tasks.

C&W have noted that understanding a sentence and following instructions are not the same thing, a point that no one would wish to dispute. They have reified this observation into a claim about separate working memories, however, and this move is an unfortunate one for a theory of cognitive processes. The motivation for the fractionated working memory is that performance in the two arenas is not particularly well correlated. By this logic, any two poorly correlated tasks should be constrained by separate working memories; the proliferation of working memories would be enormous. Indeed, it is not at all clear why C&W propose only two working memories for language; segmenting the speech stream, activating lexical semantics, parsing, and pronominal reference interpretation are very different processes, and probably abilities in these areas are not perfectly correlated, yet C&W assume that these processes are all part of one "interpretive" working memory. Thus, the decision to have one vs. two vs. twenty working memories is unprincipled.

A serious concern with all resource theories is that they are nearly impossible to falsify, because there is no theory of how reductions in resource availability will affect the myriad processes that purportedly draw on the same resource. By dividing this resource into two pools each constraining many processes, C&W have not made this concern go away; they have compounded it. C&W suggest that current data do not yield the complex interactions predicted by the single-resource theory, but in fact they review almost every conceivable pattern of data (including those with the putatively crucial interactions) and conclude that no result is inconsistent with their account. It is always possible to invent a scenario in which comprehenders allocate resources to tasks in a way that accounts for the data, especially if the sizes of the two "independent" resource pools are positively correlated, as C&W imply.

Whereas C&W go to great lengths to explain away conflicting data from studies with young normal adults, they may appear to be on firmer ground with data from patient populations. We suspect that this situation is merely an artifact of the paucity of online studies of language comprehension and working memory in these populations. Contrary to C&W's claims, our work with patients with Alzheimer's disease (AD) has shown that these patients' on-line sentence processing is impaired compared to that of controls (Almor et al. 1998). Patients appear equivalent to normal individuals in cross-modal naming only when the stimuli are constructed in such a way that subjects can ignore all but a few words in the sentence (typically the last few) and still perform the task accurately. Moreover, Almor et al. (1998) found that AD and normal subjects' accuracy producing and interpreting pronouns correlated well with performance on putatively "post-interpretive" working memory measures. Thus C&W's claim that patients with impaired "central executive functioning" have normal syntactic processing is not supported by studies that carefully manipulate the on-line processing demands of the stimuli. Instead, the ability to develop a discourse representation, a key part of "interpretive processing" in C&W's account, is related to performance on "post-interpretive" working memory tasks.

Much of the confusion in C&W's account stems from a narrow view of sentence processing and a failure to appreciate that notions of working memory are inseparable from views of processing architecture. C&W make several mistakes in their discussion of sentence-processing theories, including (1) an inconsistent blending of constraint-based and reanalysis approaches to ambiguity resolution and (2) questioning the lack of reading effects in ambiguous sentence regions for ambiguities in which no theory predicts any effects in this region. At the architectural level, C&W's account (as well as the single-resource theory) incorporates the assumption that sentence comprehension consists of building syntactic representations word by word as the basis for semantic interpretation. Working memory resources are needed for storage of partial processing results. The constraint-based account that we advocate holds that sentence comprehension involves the parallel application of multiple probabilistic constraints from sentential and discourse context. In connectionist instantiations of this view, there is no distinction between storage of linguistic knowledge, comprehension processes, and working memory resources. An individual's "capacity" to process language is realized as the efficiency of passing activation through a network and is constrained by the interaction of network architecture and experience. Including the notion of working memory resources adds nothing to this account.

Distinguishing interpretive and post-interpretive processes

Fernanda Ferreira

Department of Psychology, Michigan State University, East Lansing, MI 48824-1117. **fernanda@eyelab.msu.edu**
eyelab.msu.edu/people/ferreira/fernanda.html

Abstract: A separation between interpretive and post-interpretive processes is central to Caplan & Waters's theory of language comprehension. This commentary raises some issues that are intended to help sharpen the distinction.

Caplan & Waters (C&W) present an excellent overview of their research. Their work demonstrates that differences in working memory capacity do not affect sentence-interpretation processes themselves but may influence operations performed on the output of those processes. I find the general approach the authors take convincing, but a number of questions can be raised concerning the distinction between interpretive and post-interpretive events during sentence comprehension. Let me begin by acknowledging that C&W do go some way toward providing clarification on this point in their section 4, entitled "Discussion: Fractionating verbal working memory." There they suggest that interpretive processes include accessing words, computing prosody, assigning thematic roles to syntactic constituents, establishing coreference (although presumably not coreference relations that rely extensively on real-world plausibility), and determining a sentence's focus-presupposition structure.

What strikes me as requiring more justification are the authors' assumptions about what constitutes post-interpretive processes. Likely everyone would agree that it is not really language that tells us that if Harvey used to sell junk bonds and now he sells pencils on a street corner he likely lost money, but some of the other processes labeled as post-interpretive seem more controversial. For instance, C&W seem to assume that events involved in sentence reanalysis are post-interpretive. The examples they give include *We hated the cheap hotel room because of all the bugs we saw in it; We realized our conversations would not be private;* and *The aggressive trial lawyer questioned in minute detail by the judge hesitated.* The first presumably involves lexical reanalysis and the second syntactic reanalysis. The problem here is that certainly not all cases of either require post-interpretive processing. For example, the sentence *The team defeated in the Super Bowl vowed revenge next season* might require repair once *vowed* is encountered, but re-analysis does not seem to require more than the basic interpretive processes of the sentence-comprehension system. Indeed, recently Fodor and Inoue (in press) proposed a theory of syntactic revision that is deliberately designed to keep reanalysis internal to the sentence-interpretation mechanism. In their theory, when a word cannot be incorporated into the current syntactic phrase marker (*vowed* in the Super Bowl example above), it is "attached anyway," and then the parser faces the syntactic consequences of that attachment by moving right to left through the tree, making necessary adjustments. Therefore, it seems that C&W should say much more about what sorts of repair processes might be post-interpretive and what sorts are not.

Another question concerns C&W's assumption that sentences with more than one proposition invoke post-interpretive processing. First, I do not understand how the sentences they give as examples contrast in number of propositions; second, I do not see why number of propositions by itself should matter. The authors devote several paragraphs in the article to their argument that the processing of multi-propositional sentences interacts with resource limitations because such sentences require post-interpretive processing, but they do not really spell out how this is supposed to work. Let us take an example from the beginning of the paper that is meant to illustrate the contrast:

1. The boy hugged the girl and the baby.
2. The boy hugged the girl and kissed the baby.

According to C&W, sentence (1) is made up of just one proposition, but sentence (2) is made up of two. The idea seems to be that in sentence (1) the boy is simultaneously hugging the girl and the baby (the three are squeezed into one bear hug), but in sentence (2) the boy hugs a girl and also kisses a baby (but notice he could do both at the same time; he could be hugging both, but kissing only the baby). The claim that there is only one proposition in sentence (1) assumes that comprehenders create a particular mental model for the sentence, one in which a boy hugs both a girl and a baby at the same time. The meaning of the sentence is also compatible, though, with a situation in which the boy hugs the girl and then he hugs the baby, in which case clearly the sentence contains two propositions. It could even be argued that the sentence contains two propositions even in the simultaneous case, because the representation indicating the meaning of the sentence has to indicate somewhere what agent acted on the baby. The point here is not that the analysis assumed by C&W is wrong but rather that their position requires a set of assumptions that are vulnerable to challenge and so must be explained in more detail.

I will end by repeating that I generally like the approach the authors take in this work. Recently we have begun examining how unimpaired undergraduates interpret unambiguous sentences such as *The lawyer was sued by the doctor* and *The dentist was pulled by the tooth.* We are finding striking evidence that sentences that are implausible by virtue of world knowledge – in the authors' terminology, by post-interpretive processing – are processed quite differently from sentences whose sensibility can be ruled out based on feature mismatch (a tooth cannot be an agent because agency requires animacy). Even more intriguing is our finding that, if the two participants in an implausible event (e.g., *The lawyers were sued by the doctor*) differ on some grammatical feature (plurality in this case), confusion occurs much less often than when the noun phrases cannot be distinguished with a grammatical feature. We believe the distinguishing features allow the two noun phrases to be unambiguously bound to their appropriate thematic roles, and as a result misinterpretations are much less likely to occur (the two noun phrases cannot swap positions). Data of this sort provide independent and compelling evidence for the original and creative theoretical perspective C&W present in their target article, because they require an explanation that assumes a separation between processes that compute interpretations and so are internal to the language system, and processes that evaluate and modify those interpretations and are therefore external to the system.

Modularity, segregation, and interactions

Karl J. Friston

The Wellcome Department of Cognitive Neurology, Institute of Neurology, London WC1N 3BG, England. **k.friston@fil.ion.ucl.ac.uk**

Abstract: This commentary considers how far one can go in making inferences about functional modularity or segregation, based on the sorts of analyses used by Caplan & Waters in relation to the underlying neuronal infrastructure. Specifically, an attempt is made to relate the "functionalist" approach adopted in the target article to "neuroreductionist" perspectives on the same issue.

Caplan & Waters (C&W) provide compelling arguments in support of the notion that there is a separate working memory system for assigning the syntactic structure of a sentence by noting that there is no interaction between structural complexity and other processes that require working memory. This is a nice example of a general approach to inferring something about the cognitive architecture of a system given only its outputs (responses) to inputs or stimuli that can be manipulated over a number of dimensions. The approach is related to Sternberg's revision of Donder's subtractive method (Sternberg 1969) in which the interactions among various processes are used to infer something about their functional organization. Both depend critically on interactions, and it is this theme that will be pursued in the context of neurobiologically motivated analyses of functional anatomy. The key question, addressed by C&W, is whether the mnemonic aspects of interpretive sentence processing are modular or functionally dissociable from other (e.g., post-interpretive) aspects. This question is purely functionalist but implies the existence of distinct neuronal systems that mediate this processing. An explicit version of the same question is "Is there functional segregation, in terms of neuronal systems, for the structural complexity of sentences?" Much is invested in the term "functional segregation" here, and it is worthwhile considering what it means.

The brain appears to adhere to two principles of organization, functional segregation (Phillips et al. 1984) and the functional integration of segregated systems, such as cortical areas, subareas, neuronal populations or individual cells. "Functional segregation" refers to the selective neuronal responses to specific sensorimotor attributes. Consider the cortical area V5, a motion-sensitive area (Zeki et al. 1991) that can be characterized in terms of population responses (as measured with fMRI or local field potentials with electrode recordings) to visual stimuli. To demonstrate functional *specialization* for motion, one would have to show a high degree of mutual information between V5 responses and visual motion. For this area to be functionally *segregated,* there is a further requirement that there are no responses to other attributes, such as changes in colour. This constraint has important consequences for the context sensitivity of speed-dependent responses: imagine that we characterised the receptive field $f(s,\lambda)$ of a V5 cell as a function of stimulus speed (s) and wavelength (λ). Under segregation, the responses (x) over different speeds should be the same for any two wavelengths:

$$x = f(s,\lambda_1) = f(s,\lambda_2) = f(s).$$

In short, the speed-dependent responses would be insensitive to the colour of the stimulus. In other words, functional segregation for a specific attribute implies that the same responses will be elicited by changes in that attribute irrespective of the other stimulus attributes (or more generally the context). Imagine that we measured the response of V5 to a changes in speed using long-wavelength (red) stimuli and short-wavelength (blue) stimuli. If we found a response difference (i.e., a context-sensitive, speed-dependent response), then functional segregation for motion per se has to be rejected and we would infer that this area preferred red (or blue) motion (see Fig. 1 for a schematic illustration of the receptive fields implied by these situations). This context sensitivity is simply revealed by the interaction between speed and wavelength in predicting the response. Precisely the same construct is used by C&W. In their formulation the response is in terms of speed and accuracy of sentence processing and the stimulus attribute of interest is structural complexity. They have examined the complexity-dependent responses (low vs. high) in two contexts to see if there was an interaction. Failing to find an interaction allows them to assert that functional segregation is a sufficient model for the observed findings. The two contexts were low and high concurrent working memory load or subjects with low and high capacity.

The above argument suggests that functional segregation should involve a series of simple subtractions in different contexts in order to demonstrate the context-invariant nature of selective responses. In functional neuroimaging this is the tenet of "conjunction analyses" (Price & Friston 1997). A conjunction is defined as the presence of a main effect (e.g., of structural complexity) in the absence of interactions (e.g., with memory load). It would be interesting to revisit this issue with functional neuroimaging using a factorial design wherein structural complexity (high and low) was crossed with memory load (sub- and supraspan). It is possible that some brain regions would show functional segregation for complexity (as identified with a conjunction analysis). However, it is likely that some regions would show regionally specific interactions, speaking to the integration between inter-

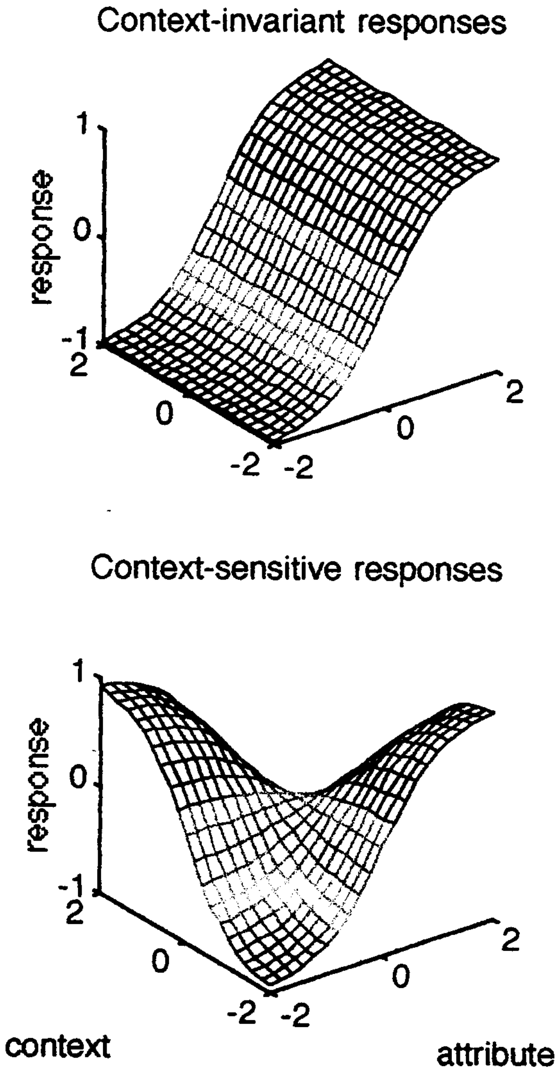


Figure 1 (Friston). Schematic to illustrate the relationship between interactions and context-sensitive responses. The upper panel shows the response of a neuronal system to changes in some attribute to which it is sensitive. In this instance the responses are insensitive to the context (i.e., there are no interactions between attribute and context. Conversely the responses in the lower panel evidence strong interactions and context-sensitivity. These response profiles can be thought of as "receptive fields."

pretive sentence processing and the mnemonic processes engaged by digit recall. Finding such effects would not detract from inferences about segregation in other brain areas but would speak to the integration, or coordination, of segregated processing systems (Friston et al. 1996). Although, to my knowledge this experiment has never been carried out, it seems extremely well justified by the psychological background provided in Caplan & Waters.

Interpretative and post-interpretative processes in sentence comprehension

Edward Gibson[a] and Rose Roberts[b]

[a]*Department of Brain and Cognitive Sciences and Department of Linguistics and Philosophy, Massachusetts Institute of Technology, Cambridge, MA 02139 and;* [b]*Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139.* **gibson@psyche.mit.edu rose@psyche.mit.edu**

Abstract: We discuss several issues raised by Caplan & Waters's distinction between interpretative and post-interpretative processes in sentence comprehension, including the nature and properties of the two systems, problems with measuring their respective capacities, and the relationship between the hypothesized separate-language-interpretation-resource (SLIR) and the general verbal working memory system that supports post-interpretive processing.

Caplan & Waters's (C&W's) proposed distinction between interpretative and post-interpretative processing leads to several interesting questions.

First, what is the nature of the interpretive system and the resources available to it? For example, what is it about the interpretive system that makes an object-extracted relative clause, as in sentence (1), harder to comprehend than a subject-extracted relative clause, as in sentence (2)?

1. The woman whom the waiter insulted left the room.
2. The woman who insulted the waiter left the room.

The range of constructions that have been tested thus far with respect to C&W's claims about the working memory resources required for interpretative processing is quite narrow. If the observed performance differences between aphasics and control groups on these constructions is due to differences in available interpretative resources in the participant groups, then the same difference should be present with respect to other sets of linguistic constructions that differ in the quantity of interpretative resources that they require in their comprehension. To test these kinds of predictions of C&W's claims, it is necessary to have a theory of the resources required in sentence interpretation. One recent theory of the computational resources used in sentence processing is provided by Gibson (1998). According to this theory, the complexity of a sentence is heavily affected by the distance between words and their dependent constituents in the sentence. Distance is measured in terms of the number of previously unmentioned discourse referents separating a word and its dependent constituent, where a discourse referent is an object or an event. Thus the complexity difference between sentence (1) and sentence (2) is caused by the fact that in sentence (2) the relative pronoun "that" is immediately adjacent to the verb on which it depends ("insulted"), whereas in sentence (2) the new discourse referent "the waiter" intervenes between "that" and "insulted." Given a theory of interpretative resources, other sentence structures can be compared with respect to their on-line processing complexity. Having an explicit interpretative complexity theory allows one to make and test predictions about complexity differences across constructions within a language and across constructions in other languages.

Second, what are the properties of the working memory system that is used for post-interpretive processing, and what is the best way to measure this system's capacity? As C&W point out, researchers studying human cognition use several different tasks to assess working memory capacity. The implicit assumption is that these tasks are all relatively interchangeable measures of the same underlying function and, thus, that results from experiments using different tasks can be compared directly. C&W report the results of an experiment showing that some common working memory measures are neither highly intercorrelated nor highly reliable. In a recent experiment, Roberts and Corkin (1998) have extended this result: they found that commonly used measures of working memory, including reading span, backward digit span, and N-back, are not significantly intercorrelated and are sensitive to different predictor variables.

C&W suggest that the reading span measure used by psycholinguists to assess working memory capacity measures the post-interpretive system used for operating on parsed input, rather than the interpretive system, and that individual differences observed in this measure are due to differences in the capacity or efficiency of the post-interpretive system, not to individual differences in interpretive processing. They state that the limitation demonstrated by low-capacity subjects is one of "retaining in memory information about the propositional content of

https://doi.org/10.1017/S0140525X99321786 Published online by Cambridge University Press

a sentence." We have observed just such an effect in an experiment measuring response accuracy to probe questions about sentences containing two to five propositions (Roberts et al. 1997). When subjects are divided into a high-capacity group and a low-capacity group using the reading span measure, both groups are near ceiling with two-clause sentences (99% vs. 92%) and are equally inaccurate with four-clause sentences (34% vs. 31%). However, high-span subjects are significantly more accurate in recalling the propositional content of three-clause sentences (73% vs. 54%). These data provide evidence that the reading span measure divides subjects according to their ability to retain information about the propositional content of sentences, a post-interpretive process. There is currently no measure of the capacity of the interpretive processing system that does not tap post-interpretive processes; designing such a measure remains a goal for further research.

Third, what is the relation between the systems used for interpretive processing and the general verbal working memory system used for post-interpretive processing? C&W claim that the separate language interpretation resource (SLIR) is a separate module within working memory, but what is evidence for this claim? Given that deficit–lesion correlation studies and imaging studies point to different neurological substrates for the two systems, and that C&W have experimentally demonstrated that they do not interact, it seems more plausible to hypothesize that the SLIR and general working memory are entirely distinct. In any event, the fact that the two systems appear to be neurally distinct provides strong evidence against Just, Carpenter and colleagues' view (see, e.g., Just & Carpenter 1992) that the same units are used for interpretative processing and storage within working memory.

What do working-memory tests really measure?

Michael J. Kane,[a] Andrew R. A. Conway,[b] and Randall W. Engle[c]

[a]*Department of Psychology, Georgia State University, Atlanta, GA 30303-3083;* [b]*Department of Psychology, University of Illinois at Chicago, Chicago, IL 30332-0170; and* [c]*School of Psychology, Georgia Institute of Technology, Atlanta, GA 30332-0170.* **mkane@gsu.edu aconway@uic.edu re23@prism.gatech.edu**

Abstract: Individuals may differ in the general-attention executive component or in the subordinate domain-specific "slave" components of working memory. Tasks requiring sustained memory representations across attention shifts are reliable, valid indices of executive abilities. Measures emphasizing specific processing skills may increase reliability within restricted samples but will not reflect the attention component responsible for the broad predictive validity of span tasks.

We thank Caplan & Waters (C&W) for pointing out the weaknesses in the view that language interpretation and parsing are limited by general, domain-free working memory resources (see, e.g., Just & Carpenter 1992). Based largely on their earlier work (Caplan & Waters 1990), Engle and Conway (1998) concluded that there is little evidence to support the claim that the phonological loop or the central executive plays an important role in the early stages of language processing. We have argued that the working memory system is theoretically equivalent to a view of human intelligence proposed by Cattell (1963) and Horn (1980), which distinguishes between a domain-free general fluid intelligence and domain-specific elements of crystallized intelligence, which are peculiar to knowledge, skill, and talent.

Our view, derivative of Baddeley and Hitch (1974), sees working memory as a system of (1) activated traces, which are the output of domain-specific processors, and (2) domain-free controlled attention (Engle et al., in press b; Kane & Engle 1998). Individuals may vary in the quality and skill of their domain-specific processors. They may also vary in their general fluid ability to bring controlled attention to bear on managing information in the temporary storage buffers and on resolving conflict at output (cf. Shallice & Burgess 1993). Thus, variability on working memory tasks can reflect individual differences in the domain-specific processors, the domain-free controlled attention, or both.

We argue that the general capability to sustain an active memory representation in the face of attention shifts or distraction is what drives the relationship between working memory span measures and diverse higher order capabilities. C&W complain that the processing component of working memory span tasks "introduces a dual-task element into these tasks and allows them to be heavily affected by the capacity for divided attention." We argue, however, that the divided-attention component is the strength of working memory span tasks. The divided-attention component is critical to the broad predictive validity of span tasks (and, therefore, of fluid intelligence tests) across task and stimulus domains.

All studies that use individual or group differences to make inferences about working memory involve the measurement of constructs and the partitioning of variance in those constructs. If the range of general ability in a sample of subjects is sufficiently broad, then measures of working memory capacity, such as the reading span task (Daneman & Carpenter 1980) and the operation span task (Turner & Engle 1989), will be reliable. Furthermore, they will reflect the ability of both the domain-free central executive and the domain-specific processors necessary to perform the tasks because no working memory task is a pure test of either the central executive or the slave systems (Engle et al., in press a). If, however, a sample's range of general ability is narrow, as it often is at prestigious universities, there will be little variability in the general fluid ability construct. Therefore, the variability in performance for the restricted sample will reflect primarily differences in the skill of the domain-specific processors necessary to perform the task.

To the extent that the "processing" component is entered into the working memory span score as C&W suggest, reliability may increase. However, the construct measured by that score is more likely to reflect the domain-specific processors required for that task and not the general controlled attention–central executive construct. In fact, our own work (e.g., Conway & Engle 1996; Conway et al. 1998) suggests that processing skills within span tasks can actually suppress the relationship between working memory scores and higher order cognition. Thus, the "innovation" proposed by C&W may increase reliability in a restricted sample, but it will be a good test only if the goal is to measure specific processing skills. It will not be a good test if the goal is to measure general domain-free controlled attention capacity. If the goal is to measure the domain-free executive component of working memory, then we recommend administering a battery of different working memory tasks that share the dual-task quality mentioned above but differ in the domain-specific processes required to perform the task.

C&W questioned the reliability of the reading span task, but, as we pointed out above, low reliability can result from sampling a restricted range of individuals. Kitty Klein and William Fiss of North Carolina State University recently conducted an extensive analysis of the operation span task (Turner & Engle 1989). They tested 33 subjects at three different times, the second time 3 weeks after the first and the third time 6–7 weeks after the second. The corrected reliability was .88 and the stability scores ranged from .76 to .92.

As for the "fractionation" of working memory, we would argue that C&W's measures reflect quite specific skills at parsing and processing syntactically complex sentences. These processes occur at a pre-interpretative level and thus do not generally make demands on the central executive. The two different components they propose reflect the outputs from specific processors that do their work with little need for limited-capacity, controlled attention. For example, adding a secondary working memory load to a

https://doi.org/10.1017/S0140525X99321786 Published online by Cambridge University Press

syntactic processing task has no detrimental effect. Because syntactic processing may proceed unhindered by a load, it may be performed relatively automatically, without much controlled attention. However, secondary load tasks do interfere with syntactic processing when the load-task stimuli are interleaved with the syntax task, demonstrating that, when syntactic information must be sustained across an attention shift, it suffers significantly.

Rather than assign syntactic processing to a specialized component of working memory, then, we suggest that it operates independently of the central executive. That is, working memory capacity is needed only under attention-demanding circumstances, and, insofar as syntactic processing appears to be immune to divided-attention conditions, it likely occurs relatively automatically. Caplan and Waters (1990) argued that the phonological loop may be required in some syntactic parsing circumstances, such as when subjects are "garden-pathed" or when many words must be maintained in active memory. Why not use that interpretation for the present work?

The age invariance of working memory measures and noninvariance of producing complex syntax

Susan Kemper and Karen A. Kemtes

Department of Psychology and Gerontology Center, University of Kansas, Lawrence, KS 66045. **skemper@ukans.edu**
www.lsi.ukans.edu/clang/sklab.htm **kkemtes@ukans.edu**

Abstract: In challenging current conceptions of the role of working memory in sentence processing, Caplan & Waters consider studies comparing young and older adults on sentence processing. This commentary raises two challenges to Caplan & Waters's conclusions: first, working memory tasks appear to be age invariant. Second, the production of complex syntactic constructions appears not to be age invariant.

Caplan & Waters (C&W) raise a number of questions about the role of working memory during sentence processing. They consider evidence from studies comparing young and older adults on sentence processing tasks relevant to this issue because older adults are typically found to exhibit working memory deficits on a variety of tasks. This commentary raises two challenges to C&W's conclusions. First, working memory measures appear to be age-invariant. Second, the production of complex syntactic constructions appears not to be age-invariant.

Age invariance. C&W question whether the tasks commonly used to measure individual differences in working memory are reliable and stable. An additional concern, not considered by C&W, is whether the tasks are age invariant. As Stine (1995) notes, even when younger and older adults are matched based on working memory scores, investigators might inevitably find that older adults have processing difficulties that are seemingly unexplained by their performance on the working memory tasks.

Horn and McArdle (1992) explain that the latent variable approach to constructing measurement permits the test of two common concepts that receive little attention in intergenerational research: construct validity and construct invariance, across samples such as age groups. Horn and McArdle define measurement invariance as factorial invariance: the characteristic that measurements are composed of linear composites, that is, item scores summed to form a total score, is stable across measurement occasions of same or different samples. the authors present four criteria that are sufficient to determine factor invariance: (1) Factor patterns are equivalent across groups, a necessary condition for factor invariance to hold. (2) Factor variances are equivalent across groups. (3) Factor covariances are equivalent across groups. (4) Factor averages of composite scores are equivalent across groups. A fifth criterion they impose is whether cross-products are equivalent across groups. If the answer is "yes," then the answer to questions (1–4) is necessarily "yes"; if the answer is "no," invariance is not necessarily falsified.

To date, few studies have used a latent variable approach to assess the age invariance of working memory measures. As an example of this approach, Kemtes and Kemper (1998) assessed working memory performance for 248 adults (young and old) in four separate studies using a variety of verbal working memory tasks. With the latent variable approach, they tested whether a simple within-studies factor analytic structure with a single latent variable of working memory was reliable across age groups. For three of the studies, this single-latent-factor working memory model was validated across age groups, suggesting that a common construct was measured. Kemtes and Kemper's finding of factor invariance across these different samples strongly supports the conclusions that working memory, as measured by the digit spans, reading span, and listening span, is age invariant.

Age noninvariance. C&W also review the existing literature on the effects of aging on sentence processing and conclude that there are few age differences in on-line sentence processing although post-interpretative processes of question answering and text recall may be affected by aging. However, they overlook an additional body of research on age differences in production. Older adults show a reduction in their production of complex syntactic constructions such as those involving subordinate and embedded clauses (Kemper 1987; 1988; Kemper et al. 1989; 1992; Kynette & Kemper 1986). The age-related decline in syntactic production is somewhat greater for left-branching constructions (e.g., *The gal who runs a nursery school for our church is awfully young*) than for coordinate or right-branching constructions (e.g., *She's awfully young to be running a nursery school for our children*). This asymmetry in the production of right-branching constructions versus left-branching constructions provides strong evidence for the effects of working memory limitations. Left-branching constructions, including center-embedded object relative clauses such as *The dog that the man that the cat bit chased escaped,* are typically considered more complex than right-branching constructions such as *The cat bit the man that chased the dog that escaped* (Gibson et al. 1996; Lewis 1996b).

Cheung and Kemper (1992) investigated the relationship between age, working memory, and production using a number of different ways of measuring linguistic complexity including: mean length of utterance (MLU; Chapman & Miller 1984), developmental sentence scoring (DSS; Lee 1974), developmental level (DLEVEL; Rosenberg & Abbeduto 1987), two alternative ways of measuring Yngve depth (Yngve 1960), and two variants of Frazier's (1985) node count. In addition, propositional density (PDENSITY), based on Kintsch and Keenan's (1973) analyses of text difficulty, was computed in order to assess whether semantic content covaries with grammatical complexity. Cheung and Kemper (1993) applied structural equation modeling to these linguistic complexity metrics using language samples from younger and older adults. The best-fitting model fit the data by specifying two correlated factors, verbal ability and working memory. Age was negatively associated with working memory, leading to a decline in digit span with advancing age, and was somewhat positively associated with verbal ability, reflecting a slight improvement in vocabulary with advancing age. Working memory was related to three syntactic factors: length, measured by MLU; the amount of embedding, measured by MCU; and the type of embedding, measured by DSS and DLEVEL as well as by both Yngve depth metrics and both Frazier counts. Finally, verbal ability predicted another linguistic factor, semantic content, measured by PDENSITY, which was not correlated with the syntactic factors.

Working memory limitations associated with aging affect older adults' production of complex syntactic structures. A key determinant of syntactic complexity, affecting the DSS, DLEVEL, Yngve, and Frazier metrics, is whether embeddings occur in the main-clause subject of left-branching constructions such as "*Going to the St. Louis World Fair* was a major undertaking" or in the

https://doi.org/10.1017/S0140525X99321786 Published online by Cambridge University Press

main clause predicate of right-branching constructions such as "I enjoyed *going to the St. Louis World Fair.*" The ability to produce left-branching constructions appears to be especially vulnerable to aging owing to working memory limitations.

ACKNOWLEDGEMENTS
Preparation of this commentary and supporting research were supported by the following grants from the National Institute on Aging: RO1AG06319, RO1AG09952, KO4AG0043, and T32AG00226.

Aphasia, prefrontal dysfunction, and the use of word-order strategies

Herman H. H. J. Kolk and Robert J. Hartsuiker

Nijmegen Institute for Cognition and Information, Catholic University Nijmegen, 6500 HE Nijmegen, The Netherlands. **kolk@nici.kun.nl; hartsuiker@nici.kun.nl www.nici.kun.nl**

Abstract: Caplan & Waters's neuropsychological evidence for two types of verbal working memory rests entirely on a very restricted definition of "syntactic complexity," one in terms of word order. This opens the possibility that the dissociation they observe relates to the differential use of word-order strategies rather than to the structure of verbal working memory.

Caplan & Waters (C&W) present a substantial amount of neuropsychological data obtained in sentence comprehension tasks. These data appear to demonstrate the existence of a dissociation. On the one hand, there are aphasic patients who show a negative effect of "syntactic complexity." On the other hand, we have the patients who are assumed to suffer from prefrontal dysfunction – Parkinson and Alzheimer patients – for which there is an effect of the "number of propositions" but not of "syntactic complexity." C&W take this dissociation to indicate the existence of two types of verbal working memory. It is this interpretation that we wish to challenge.

1. *The argumentation crucially hinges on the effect of word order.* Two effects are discussed. First is the effect of number of propositions. In this respect, the two groups do not differ: both show a negative effect. Second is the effect of "syntactic complexity." "Syntactic complexity" is defined primarily in terms of word order. Word order can be canonical or noncanonical. With aphasics, sentences with canonical word order lead to better performance than sentences with noncanonical word order. For Alzheimer and Parkinson patients, there is no difference.

2. *Word-order strategies make performance on canonical sentences better than on noncanonical sentences. It has been widely assumed that aphasic patients use such strategies.* If patients, as a consequence of a verbal working memory deficit, lose track of the syntactic representation, they could compensate for this state of affairs by choosing thematic roles directly on the basis of word order. In particular, they could associate the agent role with the first noun phrase (NP) of the sentence (cf. Caplan 1983; Caplan & Futter 1986; Caplan & Hildebrandt 1988; Grodzinsky 1990). It is not necessary to assume that patients use such a strategy invariably. It is also conceivable that processing failure occurs intermittently and that the strategy is invoked only in case of processing failure (cf. Kolk & Weijts 1996). If so, every time a patient would use it, the response would be correct with canonical and incorrect with noncanonical sentences.

3. *There is no direct evidence that noncanonicity leads to a higher processing load in aphasic patients.* The idea that noncanonicity represents an independent load factor is attractive, in particular because it relates to argument movement in linguistic theory. Nevertheless, an experiment showing the idea to be valid, in which the strategic factor is controlled for, is still missing. There is ample room for doubt. In a production task, in which Dutch-speaking agrammatics had to order constituents written on cards, we have compared subject-verb-object (SVO), subject-object-verb (SOV), and verb-subject-object (VSO) orders. Generative linguists consider Dutch to be an SOV language (Koster 1975). When word order was varied within a single clause, there was little difference between the three orders. When word order was varied within an embedded clause, all three word orders became significantly more difficult, with some advantage for the SVO order (Kolk & van Grunsven 1985). This result indicates that, at least in production, it is the phrase structure complexity rather than the canonicity that determines computational load.

4. *The strategy hypothesis can account for the canonicity effect that C&W observed in aphasics just as well as the syntactic-complexity hypothesis, perhaps even better.* In view of the above, it is rather surprising that the strategic factor is not at all considered by C&W. It is obvious that the complexity hypothesis makes almost the same predictions as the strategy hypothesis. The strategy in this case would be to take the first NP of the sentence as the agent of the relevant verb (the verb for which the roles are reversed on the distractor picture). There is one sentence type for which the predictions are different: the so-called object-subject sentences (e.g., "The horse kicked the elephant that touched the dog"; the distractor picture presents a horse kicking an elephant, as well as touching a dog). For this sentence type, the complexity hypothesis predicts relatively good performance, because word order is canonical. The strategy hypothesis predicts bad performance because the horse, the first NP, is taken as the agent of "touch." The strategy hypothesis turns out to make the right prediction. The object-subject sentences are about as hard as their noncanonical controls ("subject-object") and substantially harder than another canonical sentence type ("conjoined"). Unfortunately, no statistical test of the latter comparison is reported (see Fig. 1 in Caplan & Waters 1996).

5. *The absence of word-order strategies in Alzheimer and Parkinson patients could be related to a deficit in strategy generation and/or realization.* In our view, both aphasic and Parkinson/Alzheimer patients would suffer from a deficit in the general type of verbal working memory described by Just and Carpenter (1992), although the aphasics' deficit would be more severe. This assumption accounts for the presence of a proposition effect in all three groups. The absence of word-order strategies in Parkinson/Alzheimer patients would be related to another aspect of their prefrontal dysfunction and would lie in the domain of supervisory processes. In a recent paper on the supervisory attentional system, Shallice and Burgess (1996) identify "strategy generation" as an important component of this system. It could be this component itself or the realization of strategies – requiring the intactness of various other components – that is defective in Alzheimer and Parkinson patients.

Is it timing after all?

Sonja A. Kotz and D. Yves von Cramon

Max-Planck-Institute of Cognitive Neuroscience, D-04103 Leipzig, Germany. **kotz@cns.mpg.de**

Abstract: Even though there is ample evidence from the sentence-comprehension literature for specialized working memory systems in normal and patient populations, some open questions remain. One of them is an explanation for a missing "post-interpretive" processing deficit in a variety of accuracy-judgment tasks in an aphasic patient with a severe verbal working memory problem.

The target article by Caplan & Waters (C&W) provides an excellent overview of the specific effects of working memory capacities on different levels of sentence comprehension in normal and patient population studies. However, their conclusions about non-aphasic patients (dementia of the Alzheimer type [DAT] and Parkinson disease [PD] patients) with severe verbal working memory problems should at least be extended or discussed in a different context. In particular, the question arises of whether sentence comprehension deficits in aphasic or central nervous system (CNS) dysfunction patients are solely a consequence of a specific working memory resource deficit and/or a computational deficit. In other words, can there be an independent resource deficit without some interference from a computational deficit at the "post-interpretive" processing level? We will present data from an aphasic patient with verbal working memory deficits to broaden this discussion.

Patient H. G., a 45-year-old male, suffered an ischemic stroke in 1996. T2-weighted images in the axial plane taken about half a year after the stroke revealed an extended lesion in the left temporoparietal region and a small lesion in the left posterior frontal area. The temporoparietal lesion encloses (1) the left Heschl gyrus, (2) the posterior part of the superior temporal gyrus, (3) the supramarginal gyrus, and (4) the occipital gyri. The frontal lesion involves the middle portion of the praecentral gyrus and the posterior part of the frontal gyrus (F2p). The inferior frontal gyrus (F3) is spared. The functional deficit due to the frontal lesion could be twofold: (1) an articulatory deficit due to the lesion in the praecentral gyrus or (2) a verbal working memory deficit due to the lesion in the posterior F2 area. The lesioned supramarginal gyrus in the temporoparietal region should crucially affect the phonological short-term memory (STM; see Warrington 1975), with intact rehearsal processes (see unaffected F3 region). In terms of the functional–neuroanatomical correlates, H. G. should show the classical picture of an aphasic patient with an STM deficit and potentially a syntactic processing deficit during sentence comprehension.

In a first step, we tested H. G. for deficits in the articulatory rehearsal and phonological storage system. Results indicate no STM deficit in H. G. Tests for verbal working memory capacities such as verbal word span (span: 2–3) and an adapted aphasic reading span task (span: 1) as well as digit spans (forward and backward), however, indicate a severe verbal working memory deficit in H. G. Next, H. G. was tested with several sentence-comprehension paradigms engaging syntactic complexities (word category violations, subject–object relative clauses, subordinate clause violations). Furthermore, a combination of syntactic complexity with increasing propositional content was applied. We used both sentence picture-matching tasks and auditory and visual comprehension tasks in the two test batteries. Results indicate that H. G. did not show a significant deficit in sentence comprehension at the "interpretive" or "post-interpretive" level of processing. Thus, if working memory impairments exclusively influence "post-interpretive" processing, as proposed by C&W (sects. 3.3 and 3.4), H. G.'s data provide evidence against this hypothesis. Most strikingly, he did not show any effect of increasing propositional content. One example of this effect is that he scored 100% on a German version of the Token Test (DeRenzi & Vignolo 1962; Martin & Feher 1992) and in a judgment task with increasing propositional content.

To ensure that the "untimed" accuracy measures of sentence comprehension were not due to a ceiling effect, we proceeded to test H. G. in a combined syntactic and semantic judgment paradigm utilizing event-related potentials (ERPs). We assumed that if the verbal working memory deficit of H. G. has any effect on sentence comprehension it should be visible in the on-line measurements of ERP components specific to syntactic and semantic processes (see also Friederici et al., in press). Even though this judgment paradigm did not use the most complex syntactic structures nor a dual-task manipulation with additional load, H. G. displayed an interesting picture. Whereas the performance data were in the normal range, H. G. showed that on-line computation of syntactic structure and semantic integration was influenced by the input parameters. The evaluation of syntactic violations (word category errors) and semantic violations (selectional restriction errors) and correct sentences in connected speech did not elicit any of the expected ERP components in H. G. A similar picture emerged in the visual modality with a fast presentation rate (SOA: 500 msec). In a long SOA (1100 msec) visual manipulation, the N400 (onset: 350 msec) was elicited to semantic violations and a delayed P600 (onset: 1,000 msec) was elicited to syntactic violations. These results, although by no means exhaustive in the discussion of resource-capacity models, give evidence that reduced verbal working memory capacities might go hand in hand with a temporal deficit.

In-line measures of syntactic processing using event-related brain potentials

Marta Kutas and Jonathan W. King

Department of Cognitive Science, University of California, San Diego, La Jolla, CA 92093-0515. **kutas@cogsci.ucsd.edu king@cogsci.ucsd.edu**

Abstract: Scalp-recorded event-related potential (ERP) measures of reading and listening have been proved more sensitive to the time course of syntactic processing than the chronometric and behavioral data described by Caplan & Waters. ERP studies using sentences containing relative clauses indicate that there are individual differences in syntactic processing that appear at the earliest theoretically relevant time points and are attributable to working memory operations.

Leaving aside a discussion of Caplan & Water's (C&W's) specific proposal that there are two distinct verbal working memory systems, we note that C&W base much of their argument on chronometric and behavioral data from the processing of sentences containing subject-relative (SR) and object-relative (OR) clauses. The focus on the contrast between SR and OR clauses is clear; arguments concerning limitations in working memory capacity permeate the experimental literature on this topic. The decision to focus on reaction time measures, however, is less clear to us. Although a host of measures show consistent chronometric differences between the two sentences, these effects are generally observed substantially *after* the point in the sentence where they would be expected owing to a difference in working memory storage demands per se. In the following examples,

OR: The reporter who *the senator* harshly attacked admitted the error.
SR: The reporter who harshly attacked the senator admitted the error.

the noun phrase "the senator" in the OR sentence marks the beginning of the differential working memory load between the two sentence types. However, RT differences are not seen until the end of the relative clause or even later, on the main clause verb. Thus, one might question the sensitivity of the chronometric measures on which C&W base their arguments, at least if monotonic increases in memory load lead to monotonic increases in processing time, as many assume.

There is, however, one measure of brain processing that shows differential activity for OR versus SR sentences starting precisely where one would expect from the usual working memory arguments, namely, event-related potentials (ERPs). While C&W miscite who found out what and when, they do note that the processing of OR sentences is associated with a greater negativity than the processing of SR sentences. This electrophysiological pattern has been observed repeatedly during word-by-word reading in English (King & Kutas 1995) and German (Mecklinger et al. 1995; Münte et al. 1997) as well as in connected speech (Müller et al. 1997). This ERP difference between subject and object relatives is large and is more pronounced over anterior electrode sites; with visual materials, it is also left-lateralized. (Indeed, it is this topographic characteristic of the data that C&W highlighted when referring to ERP results in their section on the anatomical locus of verbal working memory processes.) But, to emphasize the point, the electrophysiological data show a much earlier point of divergence. Moreover, King and Kutas (1995) found that the

https://doi.org/10.1017/S0140525X99321786 Published online by Cambridge University Press

ERPs to the main-clause verb in both SR and OR sentences showed a greater left anterior negativity than the second verb in sentences that did not contain relative clauses. Although tempted, we will refrain from arguing from this pattern of effects that the insensitivity of the behavioral reaction time data suggests they are more likely to reflect "post-integrative" effects.

C&W also claim that there is no (behavioral) evidence of an interaction between sentence complexity and working memory capacity (WMC), and, because there is no evidence that sentence complexity has larger effects on low-WMC subjects, C&W conclude that RT data from the SR–OR contrast support their proposal for a verbal working memory subsystem specialized for syntactic analysis. Setting aside their interpretation of the behavioral data, we would like to point out that the ERP studies cited above also include notable demonstrations of individual differences in syntactic processing. In both written and spoken English, good comprehenders show large ERP differences between SR and OR sentence types, whereas poorer comprehenders show almost none (see, e.g., King & Kutas 1995; Müller et al. 1997). In good comprehenders, subject relatives elicit a frontal positivity, whereas object relatives show more negative deflections from this pattern. In poor comprehenders, the responses elicited by both the SR and the OR sentences are almost identical to each other and also identical to the response elicited by OR sentences in good comprehenders. It is likely that these good comprehenders would score higher on many (possibly language specific) WMC measures than would poor comprehenders. More sensitive measures thus *do* show the kind of individual differences C&W failed to find. Additionally, however, we find that good and poor comprehenders show ERP differences *even on simple transitive sentences* (Kutas & King 1996), data that are not predicted by any of the existing capacity theories, because these sentence types require far less capacity to process.

Whatever the ultimate subdivisions of verbal working memory turn out to be, we doubt that psycholinguists can afford to ignore the temporal information and processing perspective provided by sentence-level ERP measures. At present, although interesting and provocative, the C&W proposal awaits an exact empirical test.

Accounting for the fine structure of syntactic working memory: Similarity-based interference as a unifying principle

Richard L. Lewis

Department of Computer and Information Science and Center for Cognitive Science, The Ohio State University, Columbus, OH 43210.
rick@cis.ohio-state.edu www.cis.ohio-state.edu/~rick

Abstract: A promising approach to more refined models consistent with the Caplan & Waters hypothesis is based on similarity-based interference, a general principle that applies across working memory domains. This may explain both the fine details of syntactic working memory phenomena and the gross fractionation for which Caplan & Waters have found evidence. Detailed models of syntactic processing that embody similarity-based interference fare well cross-linguistically.

Caplan & Waters (C&W) present a compelling case for specialization within verbal working memory. They conclude their target article by looking ahead to additional specification of the hypothesis based on better developed models of working memory. In this commentary I describe some work that takes up the challenge to develop more refined working memory models. I will first discuss briefly some of the phenomena that must be accounted for by a more detailed theory. I will then argue two theoretical points: first, even though working memory may be quite specialized, there are general principles that apply across all kinds of working memory. Second, one of those general principles, similarity-based interference, may explain both the fine details of syntactic working memory phenomena and the gross fractionation and specialization for which C&W have found evidence (Lewis 1993; 1996b; 1998).

C&W start their discussion of memory demands in sentence comprehension with the notoriously difficult English double center-embedded relative clause (their sentence [1]) and go on to cite a few of the computational models and linguistic metrics that have been developed to account for the problem with such embeddings. What they did not mention is how surprisingly difficult it has been to produce models and metrics that are empirically adequate. Though it is true that there is a "remarkable degree of similarity in the measurements" that some of the models produce (sect. 1), the vast majority of these measurements do not fare well when considered against a broad range of embeddings cross-linguistically. Furthermore, there has been little independent psychological motivation for the proposed memory structures (e.g., stacks, lookahead buffers) and their associated limitations (see Lewis 1996 for a review).

As an example of the kind of empirical hurdle faced by any theory of syntactic working memory, consider a fact established by Cowper (1976) and Gibson (1991) in their seminal work: a metric based on the amount of center-embedding does not account for many difficulty contrasts in English and other languages. Consider sentences (1a) and (1b):

1a. That the food that John ordered tasted good pleased him (Cowper 1976; Gibson 1991).

1b. Der Bauer, der die Kuh, die schlechte Milch gab schlachtete ist krank.
the farmer who the cow which bad milk gave killed is sick.
"The farmer who killed the cow which gave bad milk is sick" (Hawkins 1994).

Though both constructions involve double center-embedding of sentential structures (and sentence [1b] even involves center-embedding of relative clauses), neither causes the comprehension difficulty associated with the classic example cited in the target article.

Although increasing center-embedding certainly increases difficulty, another important observation is that increasing the similarity of the embedded constituents increases difficulty, and making constituents more distinct or dissimilar in some way helps processing. This generalization is an old one; it goes back to Miller and Chomsky's (1963) original self-embedding metric and has been noted several times since (e.g., Bever 1970; Kuno 1974).

Why is this observation significant? Similarity-based interference is a principle that holds true of working memory in general. Starting with the early work of Baddeley (1966) and Conrad (1963), which identified a special system relying on phonological codes and covert rehearsal, evidence has accumulated for a wide range of distinct working memory types subject to selective, type-specific interference. The verbal versus visual-spatial distinction is the best known (Baddeley & Hitch 1974; Logie et al. 1990), but there is also evidence for distinct codes for kinesthetic memory (Williams et al. 1969), odor (Walker & John 1984), and sign language (Poizner et al. 1981), to name a few. The robust result across domains is that, when to-be-remembered items are followed by stimuli that are similar along some dimensions, the original items are more quickly forgotten (Shulman 1970; Waugh & Norman 1965).

Crucially, similarity-based interference operates within major categories as well. The most familiar within-category effect is the phonological similarity effect. Ordered recall of phonologically similar lists of words, consonants, or nonsense trigrams is worse than with dissimilar lists (Baddeley 1966; Conrad 1963; Wickelgren 1965). Related effects show up in immediate memory for American Sign Language (Poizner et al. 1981) and visual orientation (Magnussen et al. 1991).

Building on these results, and the work cited earlier by Gibson (1991), Cowper (1976), and others, I have hypothesized that similarity-based interference as a general principle applies to syntactic working memory as well. Lewis (1993; 1996b) described a

https://doi.org/10.1017/S0140525X99321786 Published online by Cambridge University Press

computational model that embodies retroactive, type-specific syntactic interference and accounts for a range of cross-linguistic data concerning difficult center embeddings. The model posited a simple buffer that could maintain no more than two constituents of a particular syntactic type.

The type specificity of the limitation is crucial to the empirical success of the model. To see why, consider again the difficult center embedding discussed in the target article, repeated here as sentence (2a):

2a. The man that the woman that the child hugged kissed laughed.
2b. The man that the woman kissed laughed.

Three noun phrases (NPs) of the same syntactic type (subjects) must be momentarily buffered in order correctly to parse the very difficult sentence (2a), whereas only two such NPs must be buffered to parse the acceptable sentence (2b). But consider now the comprehensible Japanese construction in sentence (3) below:

3. John-wa Bill-ni Mary-ga Sue-ni Bob-o syookai sita to it-ta.
John-TOP Bill-DAT Mary-NOM Sue-DAT Bob-ACC introduced say.
"John said to Bill that Mary introduced Bob to Sue" (Lewis 1993).

Such sentences do not cause the difficulty associated with sentence (2a), despite stacking up five NPs. The crucial difference, of course, is that sentence (3) requires buffering no more than two NPs of any particular syntactic function: at most two subjects, two indirect objects, and a direct object.

What this theory amounts to is adding "syntactic" to the list of immediate memory types that exhibit type-specific interference and decreased performance with increased similarity. Just as there is the well-known phonological similarity effect, there is also a "syntactic similarity effect," and one way this effect manifests itself is difficulty with center embedding. Lewis (1998) extends this theory to combine both retroactive and proactive interference into a measure of working memory load. The new model increases the empirical coverage considerably, and directly yields detailed, moment-by-moment predictions of processing load.

The similarity-based interference hypothesis can be seen as a further specification of the C&W specialization hypothesis. The good news may be that the same principle that yields fractionation at a gross level also seems to work well as a basis for quite detailed models of syntactic processing across languages. Furthermore, it should be clear that we can embrace fractionation and specialization without abandoning more general cognitive principles.

Further fractionations of verbal working memory

Randi C. Martin

Psychology Department, Rice University, Houston, TX 77251.
rmartin@rice.edu **www.ruf.rice.edu/~psyc/faculty/martin.html**

Abstract: Although the working memory capacity involved in syntactic processing may be separate from the capacity involved in word list recall, other aspects of initial sentence interpretation appear to depend on some of the same capacities tapped by span tasks. Specifically, there appears to a capacity for lexical–semantic retention involved in both sentence comprehension and span measures.

Caplan & Waters (C&W) have presented a convincing case concerning the independence of the working memory resources involved in syntactic processing and those that are tapped by various span measures such as reading span or digit span. They have amassed a large body of data showing that across different materials, experimental procedures and subject populations critical interactions are not obtained between the effects of syntactic complexity and extraneous memory load and between syntactic complexity and individuals' working memory capacity.

A less well justified claim put forward by C&W is that sentence comprehension can be divided into interpretive and post-interpretive processes and that all interpretive processes depend on capacities separate from those tapped by standard span measures. In order to make this claim they have to argue that an effect on comprehension of number of propositions (which does interact with extraneous load) occurs because the number of propositions effect is due to post-interpretive processes. However, their definition of interpretive processes includes "assigning thematic roles, focus, and other aspects of propositional and discourse-level semantics." According to this definition, it is not clear why greater post-interpretive capacity rather than greater interpretive capacity is involved in deciding the sensibility of sentences with two propositions (e.g., "The famous author wrote the play that watched the audience") versus one proposition ("It was the play that watched the audience"). If the capacities involved could be said to fall under interpretive processes, then C&W would be forced to conclude that, even within what might be termed interpretive processes, there are separable capacities devoted to the maintenance of different types of information, for example, syntactic structure versus propositions.

In several publications, we have argued that there are separable components involved in the maintenance of phonological, lexical–semantic, and syntactic information during sentence processing, and at least the phonological and lexical–semantic components play a role in short-term memory for word lists (Martin & Romani 1994; Martin et al. 1994; Martin 1995). Brain-damaged patients who have a disruption of the semantic component, but not those who have a disruption of the phonological component, have difficulty understanding sentences in which the integration of meanings of words is delayed because the meanings of a number of words must be held in an unintegrated fashion until some key word is processed (Martin & Romani 1994). For example, for sentences in which there are a number of adjectives preceding a noun ("the rusty old red pail"), the meaning of the adjectives cannot be precisely determined nor can their meaning be integrated with the noun until the noun is processed. Similarly, when a number of nouns precede a verb ("flowers, trees, and shrubs grew . . ."), the thematic role that the nouns play with respect to the verb cannot be determined until the verb is processed. These patients' difficulty in understanding such sentences is not an effect of the number of propositions, insofar as these patients could understand sentences with the same number of propositions when immediate integration was possible, such as when the adjectives followed the noun ("the pail was old, red, and rusty") or the nouns followed the verb ("the gardener grew flowers, trees, and shrubs"). It would be difficult to claim that comprehension of the sentences involving the maintenance of several unintegrated word meanings depends on "post-interpretive" processes. The ability to hold individual word meanings until they can be integrated with other words' meanings to form propositions would seem to be a fundamental ability involved in initial sentence interpretation. It should be noted that the same patients who had difficulty with the delayed integration sentences did not have difficulty making grammaticality judgments for sentences in which several words intervened between the two words signaling a grammatical error. Thus, the capacity affected was specific to maintaining lexical–semantic information and not syntactic information.

In discussing the separation between the working memory involved in post-interpretive processes and that involved in interpretive processes, C&W point to lesion localization and functional neuroimaging data. As they indicate, there is much evidence implicating the dominant perisylvian cortex in fundamental aspects of language processing. They further claim that neuroimaging studies of working memory have not implicated these areas. However, contrary to this claim, several neuroimaging studies of working memory have found activation in Broca's area and in the inferior parietal region (Brodmann's area 40; see Fiez et al. 1996 for an overview). Activation in these areas has generally been attributed to phonological retention and rehearsal. If these working memory studies used more meaningful word materials rather than letters or digits, as is typically the case, it is possible that other language areas involving lexical–semantic retention would be implicated as well (see Buckner & Petersen 1996). Thus, the lesion and neuroimaging data do not provide a clear separation between areas involved in sentence processing and those involved in standard working memory tasks. Although we agree with Caplan & Waters concerning the separation between the capacities for syntactic processing and those involved in word list recall, other aspects of sentence interpretation may in fact depend on the same brain regions implicated in working memory tasks.

ACKNOWLEDGMENT
Preparation of this manuscript was supported by National Institutes of Health grant DC-00218 to Rice University.

Backtracking? Rehearsing and replaying some old arguments about short-term memory

Rosaleen A. McCarthy[a] and E. K. Warrington[b]
[a]*Department of Experimental Psychology, Cambridge University, Cambridge CB2 3EB, England; and* [b]*Dementia Research Group, National Hospitals, London WC1N 3BG, England.* **rm107@cam.ac.uk**
e.warrington@dementia.ion.ucl.ac.uk

Abstract: We discuss the role of short-term auditory verbal storage within a working memory system. Data from single case studies of patients with left parietal lesions and selective impairment of memory span are discussed in order to address the question of the functions of short-term memory in language processing. The backup resource of auditory verbal short-term memory is required for those tasks that necessitate backtracking in order to integrate a verbal message within a developing central cognitive representation.

Caplan & Waters (C&W) present arguments for a differentiated and multifunctional verbal working memory system based on anterior and perisylvian regions of the left hemisphere. Although we are largely sympathetic to their analysis, we consider that they have given insufficient theoretical and empirical weight to the contribution of the parietal lobe and, more importantly, to the well established division of working memory into storage and executive components (see, e.g., Baddeley 1986).

C&W state that "there is strong evidence that the dominant perisylvian association cortex is critical for all language interpretative functions" (sect. 4, para. 7). This zone may constitute a core neuroanatomical substrate for "working memory" in those tasks that have high loading on executive problem-solving abilities (e.g., when disentangling complex syntax). However, C&W ignore the neuroanatomical evidence from patients with selective deficits in auditory verbal short-term storage. Damage to the inferior sectors of the left parietal lobe/angular gyrus has consistently been demonstrated in such cases (see, e.g., Della Sala & Logie 1993; McCarthy & Warrington 1990a; Warrington et al. 1971).

Do these patients have normal language processing? Yes and no (see, e.g., Shallice & Warrington 1970). Patients with dramatically reduced digit span and rapid forgetting (on the Brown-Petersen task) may have superficially well preserved language skills: their comprehension of sentences can appear to be normal. Even the earliest reports of short-term memory deficits noted that such patients were able to name objects on the basis of long spoken descriptions (e.g., "What is the name of the thin grey dust that remains after something has burned such as a cigarette?"). However, it was also apparent that their processing of nonredundant spoken information was impaired. In particular, their performance on Token Test instructions was well outside normal limits (e.g., "Put the green square on top of the red circle"; see, e.g., Warrington et al. 1971). This contrast would seem to hold the key to understanding the contribution made by auditory verbal short-term storage to the (multiply determined) task of language comprehension. Simple interpretations in terms of sheer length of message or impaired syntactic processing have been eliminated (see, e.g., McCarthy & Warrington 1990a).

Is this a language deficit, or something else? Some years ago we attacked the problem of accounting for the preserved/impaired characteristics of the short-term memory syndrome (McCarthy & Warrington 1987a; 1987b) Our cases RAN and NHB were remarkably intact on most tests of sentence comprehension (again with the exception of the Token Test), despite having reliable digit and word spans of only one item. For example, on a stringent order-dependent sentence–picture matching test, they scored close to a control level (e.g., "The clown applauded the dancer" vs. "The dancer was applauded by the clown"; Schwartz et al. 1980). This evidence, combined with the patients' remarkably well preserved ability to repeat sentences, led us to conclude that different resources were required by the span task and by many aspects of sentence processing. We considered that their on-line language system was preserved, enabling them to perform a range of linguistic operations, including syntactic analysis. However, the problem remained: What were the conditions under which their sentence comprehension failed?

Our patients were impaired when they were required to utilise information with a high propositional load (e.g., in judging comparatives such as "Which is green: a poppy or a lettuce?"). However, their problem did not appear to be exclusively linguistic. We suggested that the processes involved in understanding a sentence (rather than merely analysing its verbal content) reflected an interaction between linguistic and nonlinguistic levels of representation. When the number of propositions in a sentence was held constant but the pragmatic context was varied, the results were striking and were consistent with our hypothesis. For example, violating conventional "order of mention" rules resulted in a marked impairment (e.g., "She watered the garden after she went to the shops"). Pragmatic expectancies about the subject of a sentence and the object of an action also markedly influenced their performance. In a critical experiment, the patients were asked to manipulate one fixed and one movable doll in accordance with a spoken sentence. We found that instructions to place one doll behind (or in front of) the other could be understood only if the movable stimulus was in the subject position of the sentence.

We considered that these deficits indicated a specific role for verbal short-term memory in language understanding. Although much language comprehension can be mediated on-line, some conditions require the auditory verbal message to be reanalysed so that it can be assimilated within a developing central cognitive representation. Auditory–verbal short-term storage provides the backup facility that is used in order to backtrack and reanalyse the spoken message. Our patients were impaired under those circumstances in which sentence understanding required the use of a backup verbatim record. We considered that our evidence indicated that the following three conditions were likely to require backup and backtracking (McCarthy & Warrington 1990b):

1. When the rate of information presentation is too great for the development of a sufficiently unambiguous central cognitive representation.
2. When extralinguistic presuppositions bias the interpretation of the spoken message.
3. When the achievement of an adequate central cognitive representation requires supplementary cognitive operations to be performed on the spoken input.

These three conditions are unlikely to be mutually independent (and are certainly not an exhaustive listing). They simply exemplify those circumstances under which backtracking is required so that an utterance can be understood in terms of a central (supralinguistic) cognitive framework. We would speculate that the dynamic interplay between on-line language analysis, short-term storage, and a developing cognitive framework is partly under the control of executive and problem-solving systems of the frontal lobes, but it also depends on the contributions from widely distributed processors in both hemispheres of the brain.

C&W emphasise the complexities of "working memory" organisation without articulating the architecture or characterising the contributions made by different subcomponents of the system. Consequently, the account they give of the various short-term memory systems and their relationship to problem solving and language comprehension remains opaque. To replace one problem with another is not a solution: it is two problems! As demonstrated by C&W's overview, large-scale surveys have a propensity to generate data in which elegant dissociations and statistical artefacts are difficult to disentangle. In our view, theoretically motivated studies of single cases should be given substantially more weight; they provide a very direct route to understanding the components of working memory and other cognitive skills.

Good interactions are hard to find

Akira Miyake, Michael J. Emerson, and Naomi P. Friedman
Department of Psychology, University of Colorado at Boulder, Boulder, CO 80309-0345. **miyake@psych.colorado.edu**
psych-www.colorado.edu/faculty/miyake.html
emerson@psych.colorado.edu **friedman@colorado.edu**

Abstract: Caplan & Waters's arguments for separate working memory subsystems for "interpretive" and "post-interpretive" comprehension processes do not have a solid empirical basis. The likely involvement of a separate phonological loop makes their memory-load data irrelevant to theory evaluation, and the lack of statistical power from nonoptimal experimental designs and analyses unfairly reduces the chances of detecting the relevant interactions.

Caplan and Waters (C&W) propose a new distinction between working memory subsystems used for "interpretive" and "post-interpretive" comprehension processes. C&W's main evidence for this separate-sentence-interpretation-resource (SSIR) hypothesis comes from the lack of significant interactions between variables tapping "interpretive" aspects of sentence processing (e.g., syntactic complexity) and those reflecting verbal working memory capacities (e.g., external memory loads, reading span scores). Although they discuss an impressive array of empirical results documenting the failure to detect these interactions, we find such evidence tenuous at best. Specifically, we argue that (1) interactions concerning the effects of external memory load are irrelevant to the comparison of the SSIR and the single-resource (SR) hypotheses and (2) lack of statistical power resulting from nonoptical experimental designs and analyses unfairly reduces the chances of finding the interactions of interest.

According to C&W, the finding that external memory load usually does not impair comprehension of complex sentences more than simple sentences supports the SSIR hypothesis and poses a challenge to the SR hypothesis. C&W's logic is flawed, however. Although memory load experiments were used to support a working memory theory that does not make the "interpretive" versus "post-interpretive" distinction (Just & Carpenter 1992), external memory load places demands on a separate phonological loop component of working memory, particularly when participants start repeating the to-be-maintained digit sequence before the presentation of the target sentences (see, e.g., Waters et al. 1995). Many researchers, including C&W (Caplan & Waters 1990), have argued that the phonological loop is separable from other aspects of working memory that are implicated in the processing and maintenance of linguistic information during comprehension (Gathercole & Baddeley 1993; Just & Carpenter 1992). Therefore, the lack of significant interactions is expected because the memory load manipulation does not tap the same aspect of working memory as the language complexity and/or span group variables.

As C&W suggest, external memory load also likely taxes an executive control mechanism for dual tasking. This control mechanism, however, may be separable from the working memory used for language processing because frontal-lobe patients exhibiting disorders of executive processes can demonstrate unimpaired language performance (Eslinger & Damasio 1985), and, more importantly, the correlation between reading span and various so-called executive or prefrontal tasks is usually low (Carpenter et al. 1990; Lehto 1996). To the extent that this distinction between verbal working memory and executive mechanisms is valid, the lack of significant interactions involving memory load and language complexity variables can be accommodated easily without postulating the "interpretive" versus "post-interpretive" distinction for verbal working memory.

Another line of C&W's evidence concerns studies failing to find significant interactions between span groups and language complexity variables. Although this evidence is more theoretically relevant, the lack of such interactions does not mean that the SSIR hypothesis is correct. Without enough power, interactions are generally difficult to detect, particularly when the interactions of interest are fan-shaped, rather than crossover (McClelland & Judd 1993), as is the case with C&W's analyses. Thus, statistical power should be an important consideration when the acceptance of the null hypothesis is involved (Frick 1995).

C&W acknowledge the importance of power, but claim that their experiments had enough power because they consistently found significant interactions involving the number of propositions, a language complexity variable that they link to the "post-interpretive" aspects of working memory. However, this argument is invalid because statistical power depends on the effect size of each statistical test (Cohen 1988), and, thus, the significance of one interaction has no bearing on the statistical power of another interaction (McClelland & Judd 1993). In fact, contrary to C&W's claim, we suspect that, in many cases (if not all), their failure to detect theoretically relevant interactions was due to insufficient power.

Prevalent yet nonoptimal designs and analyses, namely, the use of highly discrete measures of individual differences (i.e., reading span scores such as 2.5 and 4.0) and the creation of arbitrary span groups (i.e., high-span and low-span groups), are the major causes of low statistical power in experiments cited by C&W. When compared to more continuous ways of scoring working memory span tasks (e.g., total number of words recalled), discrete span measures have lower power because they reduce the variance by not capturing subtle differences that may exist among individuals with the same span score. More problematic is the creation of arbitrary span groups. Although such ANOVA-based group analyses may be meaningful and acceptable *as long as they achieve statistical significance* (Maxwell & Delaney 1993), these analyses treat all members within a group as identical despite variation within the actual span scores, thus reducing power (Cohen 1978; Humphreys & Fleishman 1974; Maxwell et al. 1984; Maxwell & Delaney 1993; McClelland 1997). In fact, dichotomizing a continuous variable can reduce the variance of that variable by 20–67%, with a corresponding loss of power equivalent to throwing out one- to two-thirds of the sample (Cohen 1983). This issue is particularly relevant when investigating interactions, because measurement errors in one variable resulting from arbitrary grouping are exacerbated when that variable is part of an interaction (McClelland & Judd 1993).

Many of the experiments published by C&W (see, e.g., Waters & Caplan 1996b) and discussed in the target article actually showed a suggestive trend for the interactions predicted by the SR hypothesis. C&W dismissed such trends by arguing that they did not reach the .05 level, but we suspect that these nonsignificant trends likely will reach statistical significance if they are tested in multiple regression analyses, using a continuous measure of working memory spans. This point is particularly relevant when the target interactions were tested with multiple degrees-of-freedom omnibus tests, because focused single-degree-of-freedom contrasts for specific interaction patterns are much more sensitive than the omnibus tests (Abelson & Prentice 1997; McClelland 1997; Rosenthal & Rosnow 1985).

Given the prevalent use of the group ANOVA design, it is understandable why C&W used it to support the SSIR hypothesis (in fact, one of us is also guilty in this regard; see, e.g., Miyake et al. 1994). Ironically, however, this choice of designs and analyses reduces statistical power so much that obtaining nonsignificant interactions is not much of a hurdle for the SSIR hypothesis and, hence, does not provide much support for it. Thus, until C&W provide more direct evidence, we see no compelling reason to postulate separate verbal working memory subsystems for "interpretive" and "post-interpretive" aspects of language processing.

ACKNOWLEDGMENT
We thank Gary H. McClelland for providing us with useful statistical references and feedback.

Problems with plausibility and alternatives to working memory

Neal J. Pearlmutter
Psychology Department, Northeastern University, Boston, MA 02115.
pearlmutter@neu.edu www.psych.neu.edu/vita/pearlmutter.html

Abstract: Caplan & Waters propose a dedicated linguistic working memory to handle "interpretive" language comprehension, but there are data suggesting that more general working memory capacity can predict syntactic comprehension difficulty, and their claims depend on the existence of a principled distinction between "interpretive" and "post-interpretive" processes, which seems unlikely. Other conceptions of the source of individual differences also deserve consideration, as more flexible explanations of the phenomena.

Caplan & Waters's (C&W) proposal for a dedicated linguistic working memory system raises important issues for sentence comprehension, for language comprehension more broadly, and of course for their interaction with other cognitive systems. However, there are both data and a wider range of possible theories to consider with respect to their proposal.

C&W point out (sect. 2) how a wide range of sentence processing data claimed to support a nonspecific working memory system is in fact equivocal at best. However Pearlmutter and MacDonald (1995) provided stronger data than C&W suggest (sect. 2.1): with word-by-word self-paced reading, high-capacity readers were significantly slower at "in" and at "but was" in temporarily ambiguous sentences such as sentence (1), relative to unambiguous controls such as sentence (2) in the following.

1. The soup cooked in the pot but was not ready to eat.
2. The soup bubbled in the pot but was not ready to eat.

Low-capacity readers showed no differences between the two versions, resulting in a significant Span × Ambiguity interaction. The comprehension questions in this study did not ask about the ambiguity (and the thematic role applied to the first noun in these sentences is the same regardless of how the ambiguity is resolved). Furthermore, although high-capacity readers showed marginally better comprehension, neither group had substantial comprehension difficulty (98% vs. 94%), and their overall reading rates were nearly identical (321 msec/word vs. 320 msec/word), so a speed–accuracy tradeoff explanation is unlikely. These results are problematic for the C&W proposal.

Pearlmutter and MacDonald also showed that high-capacity readers' difficulty could be traced to their consideration of the relative plausibility of various alternatives for the ambiguity. Low-capacity readers showed sensitivity to the plausibility of the unambiguous control, but they were apparently unable to consider plausibility when faced with multiple alternatives. This might provide an alternative explanation of our results for C&W: perhaps plausibility is processed post-interpretively. In fact, though, this raises a more serious theoretical concern, which is the nature of the distinction between interpretive and post-interpretive processes. C&W provide some clear examples of each, including the creation of a constituent structure, assigning thematic roles, and various earlier processes for the former and storing information in LTM and using verbal information for other tasks for the latter. However, this distinction runs into problems in dealing with plausibility. As soon as any semantic information is allowed into interpretive processes (and this will be necessary for successful thematic role assignment), it becomes impossible to exclude a range of additional information. For example, one might try to draw the line at semantic features, as C&W appear to do, permitting animacy and sentience to be considered, but not other knowledge about referent and event properties. Unfortunately, in ambiguity resolution, the influence of nonfeatural semantic information appears identical to that of featural information (cf. Trueswell et al. 1994 with Pearlmutter & MacDonald 1992 and Tabossi et al. 1994): computation of the plausibility of a particular set of thematic role assignments on the basis of general knowledge of the events and entities involved is contiguous with computation on the basis of semantic features. There is no reason to think that the situation is different in the processing of unambiguous structures. C&W also connect (introduction, para. 6) their interpretive–post-interpretive distinction to the module–central process distinction (Fodor 1983), but the same problem arises in that debate, as discussed by Marslen-Wilson and Tyler (1987; see also Boland & Cutler 1996 and MacDonald et al. 1994 for related discussion). It is thus difficult to see where to draw the line between interpretive and post-interpretive processes, and of course this also muddies the water with respect to why they ought to rely on independent resource pools.

Two further theoretical issues deserve attention. First, the question of how much to fractionate working memory must be considered in more detail. C&W select one division within a continuum of possibilities. Although the evidence may suggest a division at roughly the point in processing that they suggest, it is not clear whether that is the only division, and all interpretive processes draw on the same pool, or whether, instead, various independent subcomponents (e.g., thematic role assignment, constituent structure generation, lexical access) each have their own.

Second, C&W seem to ignore the possibility that the explanation for all of their effects is not a matter of working memory capacity at all but is instead a matter of efficiency of processes. To some extent these approaches are interchangeable: if all processes relevant to "interpretation" are allowed to be (or become) more efficient together, then effects equivalent to those resulting from variations in working memory capacity ought to result. However, an efficiency approach offers the important additional possibility that individual processes can vary independently. On this approach, a process might be as specific as access to a particular lexical item, for example, or as broad as the complete computation of a constituent structure representation for an input string (thereby incorporating the fractionation issue as well). Such an approach remains to be worked out in detail, but Pearlmutter and MacDonald (1995) discuss how relevant differences in efficiency might arise as a consequence of differences in experience with language (see also MacDonald & Christiansen 1998), so that effects ascribed to differences in working memory capacity by C&W might instead be the result of variations across individuals in the quality of encoding of different types of information.

Components of verbal working memory

Aaro Toomela and Jüri Allik

Department of Psychology, University of Tartu, Tartu, Estonia EE-2484.
toomela@psych.ut.ee

Abstract: The target article differentiates a new, syntactic component in verbal working memory. We suggest that several more components could be differentiated to make a model of working memory complete. Next, syntax is not always separable from the subject's verbal memory capacity as measured by standard working memory tasks. Finally, interference between different processes cannot be taken as evidence for the processes sharing the same resources. Interference might be a result of active mutual inhibition.

Caplan and Waters (C&W) present several pieces of convincing evidence that demonstrate specialization within working memory. Although we completely agree with the authors' general intention to doubt the existence of a universal, all-purpose working memory, we think that they are still not radical enough in pursuit of their general objective.

Working memory is usually defined by its main function, the capacity to store and manipulate information in the service of accomplishing a task. Any processing that takes time and is not completely promiscuous must preserve previously computed results and order data to carry out required computations. Thus, the needs of general processing dictate that every information-processing task and even every stage of it, should have its own memory (Allik & Bachmann 1983). Although the idea of a single central processor and store where "it all comes together" on the stage of such a Cartesian theater (Dennett 1991, p. 107) is not completely unthinkable, it is more natural to begin with an assumption that mental activity is accomplished in the brain by many separate agents, with their own specific functions and purposes, which interact with one another.

How many components are in working memory? C&W are departing from the human working memory model, according to which it consists of three main components, the central executive, the articulatory loop, and the visual–spatial sketch or scratch pad (Baddeley 1990b; 1995). Now they are proposing that there is an additional component of working memory for assigning the syntactic structure of a sentence and using that structure for determining sentence meaning. Unfortunately, the authors of this amendment to the working model do not inform their readers, explicitly at least, whether the list of working memory components is now exhausted or if it is meaningful to expect some other specializations within working memory. However, from the general undertone of the article, it is obvious that the list of working model components is not complete, both empirically and logically. We have an impression that consequences of the possible multiplication of working memory components and uncertainty about their number are not fully realized by the authors. The increasing number of specialized components is changing the basic conceptualization of working memory (Schneider 1993). It is more natural to characterize working memory architecturally not as one unitary system consisting of a certain number of components but primarily as a system of relations between different relatively independent cognitive subsystems. Words, as inherently multifaceted entities, must be analyzed and processed at different levels: phonologic, morphologic, figurative, and semantic. Luria (1974), observing forms of memory disturbances after local brain damages, distinguished four basic types of relationships in working memory corresponding to phonologic, morphologic, figurative, and semantic aspects of verbal material. In addition, he presented evidence for separate systems to determine meaning relations coded in phrases and meaning relations coded in stories.

Thus, it can be proposed that all different kinds of relationships into which a word enters are processed by separate systems. It is noteworthy that two of these basic relations – phonologic and relations of words in phrases – correspond more or less closely to the phonological loop in Baddeley's model of working memory and that syntactic subsystem of working memory differentiated in the target article, respectively. However, this also suggests that several other specialized short-duration and limited-capacity relationships within working memory could be discovered.

Is there a specialization for morphological processing? C&W propose that there is a specialized system for language interpretation, identifying it mainly with syntactic processing. However, the syntax is not primary in the logical sense, for it presupposes the definition of the fundamental elements that are combined (Allik & Help 1985). It is possible that many aspects of language manipulation, like syntax within a word or sentence, are governed by an extralinguistic part of mental competence, for example, by a general capacity for mental serialization (cf. Restle 1970). Therefore, the distinction made by C&W between properly linguistic "interpretive processing" and extralinguistic "post-interpretative processing" is not totally convincing. In analytic languages, such as English, the "interpretive process" operates mainly within the sentence. In agglutinative languages, such as Estonian, where sentences are composed from strings of units, none of which has any independent, word-like status, the "interpretative process" has to operate within a word as well.

For example, in our laboratory we found that the memory span for the same noun in different cases (there are 14 cases in Estonian) is considerably lower than the memory span for various nouns in nominative case (Nõmm 1996). The "word stem similarity" effect is analogous to the phonological similarity effect, the reduced immediate serial recall of phonologically similar words (Baddeley 1990b; 1995). Thus, processing an Estonian sentence *Kõndisime toast tuppa, tubadest tubadesse* ("We walked from a room to a room, from rooms to rooms" or ". . .room.ELATIVE room.ILLATIVE, room.PLURAL.ELATIVE room.PLURAL.ILLATIVE") cannot be entirely independent from a subject's verbal working memory capacity measured by standard working memory tasks as is suggested in the target article.

Is the logic of separating working memory components infallible? In studies of working memory it is assumed that when two processes interfere with one another they share processing resources. It is possible, however, that such an interference emerges not because the processes have to share common limited resources but because they are inhibiting each other. Recent studies on humans as well on animals have shown the importance of inhibitory processes in cognitive functions (Kapur 1996; Irle 1987; 1990). Luria (1974) proposed several years ago that when a subject is asked to memorize verbal material, phonological relations between words are usually *actively inhibited* by semantic ones, which generally dominate over the other relations. According to this view, storage and recall are always a choice between several alternatives, the preference of one type of relationship and the inhibition of other relationships between words. For example, phonological and other similarity effects can be the result of this kind of active inhibition but not the result of difficulties with discriminating between stimuli. Analogously, findings that demonstrate interference between two processes are not sufficient for claiming that the processes share processing resources. Interference might be a result of mutual inhibition.

ACKNOWLEDGMENTS

We thank Kathryn Wycoff who helped with our grammar. This work was supported by the Estonian Scientific Foundation.

The Resource King is dead! Long live the Resource King!

John N. Towse,[a] Graham J. Hitch,[b] and Una Hutton[c]
[a]*Royal Holloway University of London, Egham, Surrey, TW20 0EX United Kingdom;* [b]*Department of Psychology, University of Lancaster, Bailrigg, Lancaster, LA1 4YF United Kingdom; and* [c]*Department of Psychology, Royal Holloway University of London, Egham, Surrey, TW20 0EX United Kingdom.*
j.towse@rhbnc.ac.uk ghitch@lancaster.ac.uk u.hutton@rhbnc.ac.uk
www.pc.rhbnc.ac.uk/jt/jt.html

Abstract: Working memory span forms an important cornerstone of current accounts of cognition, and cognitive development. We describe data that challenge the conventional interpretation of span as a measure of working memory capacity. We argue that the implications of these data undermine the analysis provided by Caplan & Waters concerning the role of working memory in sentence comprehension.

The framework of working memory advocated by Baddeley and Hitch (1974; see also Hitch & Baddeley 1976) sought to embed the processes and systems responsible for short-term memory within the context of cognitive activity generally. Rather than viewing memory as an independent system – an appendage on cognition – the authors advocated a more functional approach, whereby working memory supports many aspects of cognitive activity. In these terms, the target article of Caplan & Waters (C&W) stands as a monument to the success of the working memory framework, given their view that the post-interpretational phase of sentence comprehension (though not syntactic processing itself) reflects memory processes.

C&W also offer an important challenge to the interpretation of working memory span tasks within a single-resource (SR) theory. C&W's treatment of data often cited in support of SR theory is laudable, and in the main they provide an excellent critique. Leaving aside concerns about their focus on reading span – a task requiring sentence comprehension – as the vehicle to explain sentence-comprehension processes, where does the analysis lead? Having argued that the Single Resource King is dead, in fact very little changes, for resource sharing lives on in the post-interpretative account of comprehension and enaction. Thus, C&W perpetuate the current orthodoxy (established by the pioneering analysis of Daneman & Carpenter 1980) that working memory span measures a working memory capacity shared between processing and storage functions. This is, after all, the logic driving the comparisons between groups of "high" and "low" span, pervading both SR and "separate-sentence-interpretation-resource" (SSIR) theories.

Within the area of cognitive development too, the notion of resource sharing is widespread and offers a powerful account of age-related change. For example, changes in counting span (Case et al. 1982), a task analogous to reading span, form a central plank of Case's (1985) theory of cognitive development. Nonetheless, we believe that resource sharing can be supplanted in large part by the view that working memory span is affected by time-based forgetting. Experimental manipulations of task difficulty – traditionally thought of as affecting span by altering the balance of shared resources – may be merely manipulations of the retention interval for memory items. This conclusion is based on evidence that children's counting span is equivalent when task difficulty varies while retention time is held constant (Towse & Hitch 1995). Conversely, spans are reduced when retention time is increased but task difficulty is held constant (Towse et al. 1988). This occurs across a range of ages and tasks, including counting span, operation span, and reading span (see Fig. 1). Furthermore, we found no systematic decline in processing efficiency as memory load increased, contrary to what one would expect if processing and storage competed for the same pool of resources. What seems to be important for working memory is retention interval (the extent to which tasks provide the opportunity for children to forget stimulus items), not the consumption of "processing resources."

We believe that our data on the nature of working memory span have implications for the comparison of adults with low and high span scores; there may be several differences between span groups that have nothing to do with resource-sharing capacity. Among various possibilities are the following.

1. Low- and high-span groups differ in the rate at which representations degrade; high-span adults can endure longer retention intervals.
2. High-span adults process events more quickly, reducing the amount of forgetting that takes place (indeed, we found reliable correlations between children's working memory speed and span even after partialing out age; Towse et al. 1998).
3. Speed effects may be more subtle because faster processing allows an individual to produce more computations, or engage in more complex computations, in a given time period. Thus, even when span groups do not differ in processing time, one cannot be certain that they perform the same cognitive operations. The additional cognitive possibilities enjoyed by a quick processor may be the very key that gains these lucky souls entry into the "high-span club."

These considerations suggest that C&W miss the radicalism of their own analysis, in the real extent to which resource sharing may be a chimera, and in the need to reconsider the largely untested and potentially sterile orthodoxy that any differences between span groups are due solely to their resource-sharing capacities. Meanwhile, the fallback position, that alternative paradigms such as random generation may be the saviour of working memory assessment, remains unconvincing. Random generation is an important task, but there is increasing evidence that, as with working memory span, it is a product of complex and heterogeneous skills, difficult to render into a meaningful, singular construct (Towse 1998).

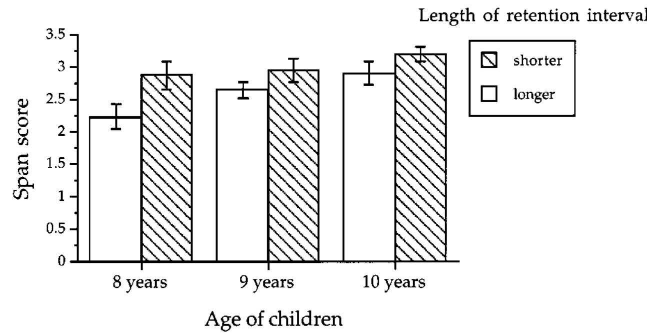


Figure 1 (Towse et al.). Reading span scores (and standard errors) adapted from Towse et al. (1998). Children read sentences and produced the missing sentence-terminal word, which they later attempted to recall. The reading time on the first and last sentences in the set was manipulated to alter the required retention interval (because memory requirements begin only when the processing product is derived) while keeping the overall work on a trial comparable. The effects of retention interval differences on span are significant for each age group.

ACKNOWLEDGMENT
This work was prepared with assistance from the Economic and Social Research Council (grant R000236113).

Sources of variability in correlating syntactic complexity and working memory

Matthew Walenski[a] and David Swinney[b]

[a]*Department of Linguistics, University of California, San Diego, La Jolla, CA 92093;* [b]*Department of Psychology, University of California, San Diego, La Jolla, CA 92093.* **mwalenski@ucsd.edu dswinney@ucsd.edu**

Abstract: Caplan & Waters's model differentiating levels of processing and the role of working memory is important and likely right. However, their claim rests on a lack of correlation between working memory and structural complexity. We examine sources of variability in these measures that remain unaccounted for (by anyone), variability that muddies a straightforward claim that the lack of correlation is cleanly established.

Caplan & Waters (C&W) have provided a review and theoretical interpretation of data that are important and in their larger picture, we believe, essentially correct. The target article provides a detailed empirical basis for a position which has had progenitor observations as early as work from the 1970s (e.g., Fodor et al. 1974), namely, that what is standardly examined in psychology as short-term memory (or working memory, or verbal working memory) may not be the same process that is involved in normal sentence comprehension (but it may be involved in task-related behavior tied to experimental paradigms in the study of language). There are important issues, however, that we believe require further clarification and delineation; some of them will inevitably blur the clear-cut distinctions C&W have attempted to draw. Specifically, previously unexamined sources of variability may contribute to the lack of a strong correlation reported between verbal working memory (VWM) and syntactic complexity, thus weakening claims based on lack of correlation.

First, we believe the distinction that C&W draw with regard to interpretive versus post-interpretive processing is correct. However, to the degree that one believes structural processing to be a coherent, individuated aspect of interpretation, VWM cannot be a sufficient mechanism for capturing the memory representations of sentence interpretation (even a VWM assigned proprietarily to language processing). It is not obvious, for example, how structural relationships could ever be captured in a representation of phonological memory. This issue somewhat recapitulates arguments underlying the move away from structuralism in linguistics; part of Chomsky's revolution was the recognition of the role of syntax in how surface form maps onto meaning. Such a mapping cannot be derived from word (phonological/morphological) forms alone (nor can it be derived solely from propositional accounts of structure). This in itself may be sufficient to account for some of the lack of covariance between memory span and syntactic complexity effects that C&W find (given that syntax is not phonological in nature and that measures of VWM do not typically attempt to take into account any unit larger than the word). VWM is essentially a claim about phonological representations. Some consideration of the nature of the processing resources that structure-building uses in interpretive level processing, as well as their relation to other types of resources (e.g., VWM), is necessary to measure accurately effects of these resources.

In addition, whereas we agree that standard VWM tests are typically indicative of post-interpretive processing, the claim of a lack of correlation of VWM with interpretive processes (structure building) has not been as clearly examined as might be desired. The evidence for a lack of correlation between VWM and syntactic complexity comes nearly entirely from tasks that present materials visually (with the exception of one auditory moving window study, to which we will return). Moreover, the reading span task itself is typically presented visually. In these cases, visual encoding and (probably) visual-to-phonological processes are involved at each stage. To dismiss VWM's role in interpretive-level processing may require a higher standard of explicitness in testing, namely, the use of verbal-only stimuli. Thus, again, some of the noise (lack of strength) in the correlations may be attributable to interference from nonverbal processes.

Despite the fact that Daneman and Carpenter (1980) find a good correlation between visual and auditory presentation of sentence memory span materials, we have found (Walenski et al. 1997) that modality of presentation does significantly affect the assessment of span. We find that sentence memory span scores are significantly higher when the test materials are presented auditorily compared to visually. Scores are probably higher for auditory-only stimuli because of the greater automaticity of auditory processes, compared to reading. Therefore, one needs to ask the following: With any reading-based task, to what extent does the mediation by material from the visual modality force the task to involve more post-interpretive (i.e., less automatic) processing than one that is purely auditory (and uses only phonological working memory)? Furthermore, rate of auditory presentation has been demonstrated to affect the structural processing (Swinney & Love 1998). Even normal but slightly slowed presentations engage structural processing (involving some degree of conscious, post-perceptual processing) different from that used for normal to fast rates of speech. This brings us back to the auditory moving window task, which uses relatively slow presentation; hence it too is likely to involve some degree of task-induced post-perceptual processing and may not reflect the more automatic comprehension processes found in normal-rate fluent auditory comprehension. Clearer tests using only materials presented in the auditory domain at normal rates are needed.

Finally, it is self-evident that to ensure as "clean" a test of these hypotheses as possible, the materials used in the assessment of span must be controlled as carefully as possible (cf. Walenski et al. 1997). The reading span materials typically used (Daneman & Carpenter 1980; Just & Carpenter 1992) have not been controlled either for syntactic complexity or for number of propositions. The fact that C&W do find better correlations for number of propositions than for structural complexity might be simply an accident related to the lack of control in the reading span materials (perhaps those materials differ less in propositional content than in structural complexity). In an uncontrolled test, more and less complex sentences will essentially be randomly distributed throughout the materials. This will affect task demands nonuniformly throughout the task and may result in low- to middle-span subjects being assigned a lower-than-true score if they encounter more complex sentences early in the task, increasing task demands for them at that point. Similarly, it may result in middle- to high-span subjects being assigned a higher-than-true score if less complex sentences are encountered late in the test, reducing task demands for them at that point. Nonstructural-complexity-equated reading span tests may accordingly provide a highly variable measure of the level of syntactic complexity that a subject can handle (concurrently with memory demands).

Again, we believe that the distinction between interpretive and post-interpretive processes that C&W draw is an important one, and that "standard" VWM is involved in post-interpretive processing. However, further work must be done to resolve variability issues resulting from the interaction of materials, tasks, and mental processes before a clean case for a lack of correlation between syntactic complexity and standard VWM can be definitely established.

ACKNOWLEDGMENT

We gratefully acknowledge support from NIH grant DC02984 for the work reported herein.

A separate language-interpretation resource: Premature fractionation?

Paul Whitney[a] and Desiree Budd[b]

[a]*Department of Psychology, Washington State University, Pullman, WA 99164-4820; and* [b]*Department of Psychology, Morehead State Unviersity, Morehead, KY 40351.* **pwhitney@mail.wsu.edu** **d.budd@morehead-st.edu**

Abstract: The target article argues for the modularity of language interpretive processes without the usual criterion that a module be informationally encapsulated. It is the encapsulation criterion, however, that gives modularity most of its testability. Without the criterion of encapsulation, testing whether relatively automatic comprehension processes use their own unique resource is a very tricky matter.

At the core of Caplan & Waters's (C&W's) argument is a question that runs as a theme through current research in psycholinguistics: What are the boundaries between the language system and the more general cognitive system? Typically, this question is framed in terms of whether the language system is composed of distinct modules or interactive subprocessors (cf. Farah 1994). Instead, C&W propose a language–cognition boundary that is based on distinctions between pools of attentional resources. As the authors point out, much of the support for the fractionation of resources comes from failures to find interactions between working memory limitations and variables related to the difficulty of sentence processing, such as syntactic complexity. C&W acknowledge that this raises an issue of the power needed to detect the interaction. We believe, however, that the authors have failed to appreciate fully the relevance of power considerations for their argument, especially if we consider power more broadly. Power, in the context used here, is not simply a matter of having enough participants but is instead a matter of whether the research methods are implemented with sufficient precision to provide a fair test of the hypothesis.

Our concerns about power stem from a key point on which single-resource and multiple-resource theorists agree. Specifically, language-interpretive processes are highly practiced and, in most contexts, relatively automatic. Fodor (1983) justified his thesis that language functions are served by modules that are functionally and neurologically separable from more general cognitive processes in part based on the relative automaticity of language processing. Similarly, C&W justify their thesis in part by the claim that the fast, overlearned nature of interpretive processes makes it reasonable to view these processes as an integrated set served by a separate resource.

It is important to note the difference that framing the issue in terms of processors or resources makes here. In the case of a modular subsystem that is fast and automatic, *and* that is informationally encapsulated, the automaticity of the system does not interfere with testing for the isolability of the module. In fact, tests of modularity have centered on the criterion of information encapsulation (e.g., Tanenhaus et al. 1995). However, if we have a separate resource used by a part of the system that operates in a relatively automatic fashion, then the automaticity of the processes compromises our ability to test the resource hypothesis. By definition, it will be difficult to detect interactions of working memory limitations with language-interpretive processes if these interpretive processes are relatively low in resource demands. Thus, it is possible that different syntactic structures take somewhat different amounts of time to compute but that the whole scale of the resource demands of syntactic parsing is low enough that tests of the single-resource view yield weak and inconsistent interactions. In fact, the patterns of results shown in Figure 4 of the target article are qualitatively similar to several similar findings reviewed by Just and Carpenter (1992). The load effects appear larger for low-span readers but, in this case, not significantly so. The power problem is exacerbated in designs such as this by treating span as a grouping variable with three levels. This approach results in a substantial loss of power compared to using only high or low spans in the design or taking a multiple regression approach with span as a continuous variable (cf. Cronbach & Snow 1977).

The problem of teasing apart different kinds of resources is further complicated by people's ability to manage working memory overloads by trading off one kind of processing for another. Of course, concerns about processing tradeoffs are at the heart of C&W's misgivings about the reading span test used by Carpenter, Just, and their colleagues. Although we agree with C&W that measures of reading span should take both word recall and sentence processing into account in computing a person's score (see also Engle et al. 1992) it is also important to realize that processing tradeoffs can take place within language tasks themselves. Data from our laboratory (see, e.g., Budd et al. 1995) have shown that low- and high-span readers sometimes tailor their processing strategies to fit their level of working memory capacity. The result is that processing by low spans may not be just quantitatively different from that of high spans but also qualitatively different. Such qualitative differences in task performance have important implications for testing predictions about interactions among reading span, load manipulations, and language-processing tasks. For example, King and Just (1991) obtained evidence that low-span readers may increase use of pragmatic information to aid sentence parsing rather than rely on syntactic information alone. With such tradeoffs occurring, a comparison of sentence reading times across conditions may be misleading if it is based on the assumption that two groups of readers are processing in a similar manner. Further, it may be that readers' performance of a given task will differ qualitatively when processing load is manipulated, as in the studies that provide digit load conditions. Unless we examine multiple aspects of the comprehension task, we might miss where participants are sacrificing text processing to maintain digits in working memory.

Perhaps a way to provide a more powerful test of the hypothesis proffered by C&W is to examine comprehension in richer contexts. Their multiple resource hypothesis predicts that tradeoffs between interpretive and post-interpretive processes should not be obtained because these processes are served by different resources. It may be possible to test this hypothesis by manipulating text processing load in a way that affects post-interpretive processing and looking for tradeoffs at more local, sentence-interpretation levels. This requires that investigators look beyond the sentence level to the level of discourse, where the ability to manipulate processing loads is less constrained than at the sentence level. In these richer contexts the SLIR hypotheses may yet prove testable, despite the relative automaticity of many sentence-interpretation processes. At present, however, we regard strong conclusions about separate language resources as premature.

Working memory and sentence comprehension: Whose burden of proof?

Arthur Wingfield

Department of Psychology and Volen National Center for Complex Systems, Brandeis University, Waltham, MA 02254-9110.
wingfield@volen.brandeis.edu **www.ccs.brandeis.edu:80/~kopp/**

Abstract: Caplan & Waters argue that the processing resources used for sentence comprehension are not drawn from an undifferentiated verbal working memory resource. This commentary cites data from normal aging to support this position. Still lacking in theory development is a specification of the transient memory representations necessary for interpretive and post-interpretive operations.

Single-resource (SR) models have appeared with striking regularity in cognitive theory, from Titchener's (1908) postulate of the need to divide limited "psychic energy" among multiple perceptual inputs to Kahneman's (1973) influential model of attentional

https://doi.org/10.1017/S0140525X99321786 Published online by Cambridge University Press

resources and their allocation. The popularity of SR models has rested partly, one presumes, on an appeal to parsimony: Why argue for separate resource systems if one will do? There also seems to be a remarkably seductive appeal to seeing cognitive resources as a sort of metaphorical fluid reservoir in a hydraulic system in which, as effort is expended on one mental operation, less effort is available for others. Within the domain of language comprehension, this SR view is echoed in the widely held position that verbal working memory represents a limited-capacity memory and resource system that carries, and hence constrains, language comprehension (Carpenter et al. 1994).

Caplan & Waters (C&W) make the case that sentence comprehension must be carried by a separate sentence-interpretation resource that is not measured by standard verbal working memory span tests. It is interesting to ask why such a multiple-resource claim represents a minority position that has to be argued, rather than placing the heavier burden of proof on the original SR theories. We can see this in the hurdles that C&W face with their arguments for a separate language-interpretation resource. Among these is C&W's unenviable position of having to prove the negative, in which they mut show that concurrent activities do not interact with syntactic complexity in dual-task settings, and they must show an absence of consistent correlations between working memory capacity as measured by traditional span tests and tests of sentence comprehension. Such correlations have been reported (Carpenter et al. 1994; Daneman & Merikle 1996). C&W argue, however, that such correlations are typically modest ones and that when they do appear they reflect post-interpretive processes rather than constraints on sentence comprehension itself.

The strongest arguments against SR models come from reports by C&W and others of neurological patients who show reduced capacities of short-term verbal memory or executive function yet show good sentence comprehension, even for sentences with quite complex syntactic constructions. Strong evidence can also be found in studies of normal aging. Although there are wide individual differences in rate and extent, age-related declines in working memory are a virtual hallmark of the aging process. Indeed, given the neural changes in the aging brain (Raz et al. 1998) it would be surprising if it were otherwise. The value of studies of language comprehension in healthy aging is that losses in processing speed and efficiency are not accompanied by a significant loss in linguistic knowledge (Light 1990). Thus, whatever differences one sees in language comprehension would be more likely attributable to working memory or other processing constraints than to a disruption in the richness of the linguistic knowledge base.

C&W recognize that elderly adults can show performance decrements with especially complex sentences, but their claim is that these deficits lie not in immediate sentence comprehension but in what they refer to as post-interpretive operations. Our own review of the literature has also revealed many published cases in which age differences in speech comprehension or recall were not increased by the need to perform a concurrent activity even when drawn from the verbal domain (Tun & Wingfield 1993).

C&W's alternative to SR theory is not yet fully formed. An important area not addressed in the target article is the question of the nature of temporary memory representations that must be involved in sentence comprehension. One example relates to acoustic–phonetic conversion leading to lexical access, a function C&W include in the integrated process that they believe is governed by their sentence-interpretation resource. The fact that speech perception is inherently context dependent demands the necessity of a syllable-sized perceptual buffer to allow for syllable identification (Mattys 1997). At a higher level has been an argument for a conceptual short-term memory (Potter 1993) that stores what might arguably be the product of C&W's interpretive operations: a transient memory representation that may supply a needed base for post-interpretive operations. C&W's call for development of their model might well be targeted to the nature of these representations and whether they are constrained by, or independent of, the postulated sentence-interpretation resource. Although the position staked out by C&W is an incomplete one at this stage, I believe it is an important step in the right direction.

ACKNOWLEDGMENTS
The author's research is supported by National Institutes of Health grant R37 AG04517 from the National Institute on Aging. Support from the W. M. Keck Foundation is also gratefully acknowledged.

Authors' Response

Issues regarding general and domain-specific resources

David Caplan[a] and Gloria Waters[b]
[a]*Neuropsychology Laboratory, Department of Neurology, Massachusetts General Hospital, Boston, MA 02114;* [b]*Department of Communication Disorders, Boston University, Boston, MA 02215.*
caplan@helix.mgh.harvard.edu gwaters@bu.edu

Abstract: Commentaries on our target article raise further questions about the validity of an undifferentiated central executive that supplies resources to all verbal tasks. Working memory tasks are more likely to measure divided attention capacities and the efficiency of performing tasks within specific domains than a shared resource pool. In our response to the commentaries, we review and further expand upon empirical findings that relate performance on working memory tasks to sentence processing, concluding that our view that the two are not strongly related remains viable in light of the material presented in the commentaries. We suggest that a productive research enterprise would be to develop the concept of working memory as a pool of resources in relation to specific tasks.

We greatly appreciate the time and effort the commentators on our target article have spent on their contributions. We have benefited enormously from these comments on our work; we hope that the airing of issues concerning this topic is of use to others as well. We also wish to apologize to any commentators who, after reading our Response, believe we have misunderstood their work. We have done our best to understand and respond to our colleagues' thoughts about our target article.

The commentaries address a wide range of topics. We recognize that we cannot respond to every comment; we have tried to deal with what we take to be major issues that the commentaries raise. We have structured this Response along the same lines as our target article. We deal first with the concept of working memory, then with the distinction we wish to draw between interpretive and post-interpretive processing, with the relationship between interpretive and post-interpretive processes and working memory, and finally with neural issues.

R1. The concept of working memory

In our target article, we used the term "working memory" to refer to a putative component of short-duration memory that performs computations on relatively small amounts of information to accomplish a task. This involves the hypoth-

https://doi.org/10.1017/S0140525X99321786 Published online by Cambridge University Press

esis that the informational storage and computational demands of many cognitive tasks draw on "resources" in such a working memory system and that limitations in these resources in part affect the throughput of such tasks. We believe that this is the sense of the concept "working memory" articulated by Baddeley and Hitch (1974). In further elaboration of this concept by Baddeley and colleagues, a "central executive" was hypothesized to be responsible for both computational and storage capacities and "slave systems" for additional storage of verbal and nonverbal representations. Some of the commentaries on our target article challenge this concept.

Kane et al. consider working memory "a system of (1) activated traces, which are the output of domain-specific processors, and (2) domain-free controlled attention." The domain-specific processors may be subject to resource limitations (this depends on how automatic they are), but the attentional system does not appear to participate in computations; rather, it schedules domain-specific processors. With respect to tests of working memory, Kane et al. say that "the divided-attention component is the strength of the working memory span tasks. . . . [It] is critical to [their] broad predictive validity." **Towse et al.** present a similar view. They argue that variation in working memory span is better explained by variation in retention interval than by variation in the degree to which a limited resource must be shared across two tasks (Towse & Hitch 1995). They therefore suggest that there is no justification from performance on such tasks for postulating a short-duration memory system in which the processing and storage demands of a task compete for a limited resource. Towse and Hitch (1995) indicated that the children in their experiments "switched" rather than "shared" attention across the counting and recall components of the counting span task. These commentators appear to think of working memory as consisting of an executive component that schedules domain-specific processing and the domain-specific processors themselves.

We are sympathetic to this perspective. In previous work, we have examined the relationship between the storage and processing components of the sentence span task (Waters & Caplan 1996c). If there is a single working memory resource, performance on the recall and sentence measures should be positively correlated at list lengths below span, because subjects with greater capacities will accomplish both the sentence and the span tasks better than subjects with lesser capacities. At list lengths above subjects' spans, performance on the sentence and recall components should be negatively correlated, because subjects' capacity limitations will force them to trade off between the two tasks. Neither of these predictions was borne out by our data, suggesting that the availability and limitation of a single working memory resource is not a major determinant of the relationship between performance on the processing and recall portions of this task. More generally, we note the analysis of resources provided by Navon (1984), cited by **Christiansen & MacDonald,** which makes clear the difficulty of providing strong evidence that dual task decrements should be attributed to shared resources. If the "more radical" view (**Towse et al.**) that there is no central resource is correct, our separate language interpretation resource (SLIR) hypothesis is trivially true, as long as language comprehension (1) is a separate domain of cognitive functioning and (2) requires processing resources.

We want to distinguish between the claims that different tasks share a single resource and that performance on individual tasks is limited by resource availability. **Christiansen & MacDonald** dispute the latter claim. For them, the concept of resources is a "theoretical soup stone" (Navon 1984) sold to an ignorant Russian peasant (us psychologists) to be put in boiling water to make soup; however, the soup's flavor depends entirely upon its other ingredients. Focusing on language, Christiansen & MacDonald argue that language processing consists of "passing activation" through a connectionist network in which there is "no distinction between storage of linguistic knowledge, comprehension processes, and working memory resources." We do not accept this view. Navon (1984) did not deal with models of performance on single tasks; in fact, his arguments assume resource consumption by individual tasks. Nor do connectionist models of individual tasks eliminate the notion of processing resources. Connectionist models have features that determine the efficiency of their operation that are independent of the specific representations or the elementary operations in a model, such as the number of hidden units in a net (Seidenberg & McCelland 1989), the number of feedback loops in the system (Tabor et al. 1997), or a built-in decay of weight strengths (Plaut et al. 1996). These features have limiting effects on computations, and, at a sufficiently abstract level, are analogous to limitations on computations imposed by other mechanisms in procedure-based models. Connectionist models also have features that serve as metrics of the complexity of an operation, such as the number of iterations through a net before a response is selected. Thus they also provide a means of measuring processing load. We believe that the concept of processing resource or working memory is not eliminated in connectionist models so much as transformed because of the nature of the models. Although the existence of a shared central resource has been questioned, the facts of complexity and floor effects in individual tasks suggest that resource utilization and resource limitations are determinants of performance in individual tasks.

R2. Interpretive and post-interpretive processing

We distinguish between processes that assign the meaning of a sentence in a discourse and those that operate on this meaning to perform other tasks. Although the distinction between interpretive and post-interpretive processing is intuitively clear in selected examples, as many commentators acknowledged, the boundary between the two types of processing and the exact characterization of each raise a variety of issues.

In our target article, we considered the interpretive stage of language processing to consist of the computation of a set of linguistic representations (lexical items, syntactic structures, intonational contours, thematic roles, focus, topic, etc.). **Pearlmutter** argues that general knowledge of events enters into the computation of thematic roles and that the distinction between linguistic and nonlinguistic processing therefore cannot be maintained. Our response to this comment is twofold: first, the evidence for the lack of informational encapsulation in language-comprehension processes is not definitive (see Frazier & Clifton, 1996, for discussion). Second, even if we grant that language processing is not as narrowly informationally encapsulated as Fodor (1983) envisaged, broadening the input into the computational processes does not render the notion of interpretive processing incoherent. If judgments of event probability and lexical statistical facts are integrated into on-line comprehension, they nonetheless enter into computations that yield the representational types we associate with language interpretation. Consider the equivalent process in vision. Many different basic visual processes can be affected by context, as visual illusions document. This does not imply that the visual system does not yield structural descriptions of objects, or face identity, or other endpoint representations that can then be utilized by other systems.

Bánréti points out that object relativization in English serves purposes such as expression of new information, topicalization, emphasis, and so on. He therefore suggests that difficulties subjects have with sentences with this construction may reflect the interface between language and the mind. We agree that object relativization in English, and many other syntactic structures that increase complexity at the syntactic level, are used to convey features of discourse structure, and we also agree that the boundary between "linguistic" and "cognitive" processes as applied to both the structure and the processing of discourse is hard to draw. We think that features such as those Bánréti mentioned – novelty of information, topic, focus, agency – are always computed as part of discourse processing and are part of the interpretive process. On the other hand, we doubt that it is the fact that object relativization expresses an unusual focus or topic that leads to difficulty with this construction in the experiments that have been performed in many laboratories. This is a testable hypothesis: if inappropriateness at the discourse level is the reason for subjects' difficulty with these constructions, the difficulty should disappear in the appropriate context.

Ferreira asks which repair processes are part of the interpretive system and which are post-interpretive, pointing out that there are models in which reanalysis is accomplished entirely intralinguistically. We did not mean that all reanalysis is post-interpretive; we believe that some reanalysis is part of the interpretation process. Reanalysis that shifts into a controlled, conscious mode is certainly post-interpretive by our criteria. The criteria for identifying this shift are admittedly subjective, but the shift is used to generate data by some researchers who attempt to distinguish between different reanalysis mechanisms (Sturt & Crocker, 1996, and references cited therein). We note that the problem raised by Ferreira is not confined by reanalysis of ambiguous structures. At some point, the processing of unambiguous, multiply center-embedded relative clauses ceases to be accomplished by unconscious dedicated routines and begins to be accomplished by problem-solving mechanisms. Metrics of structural complexity (see below) provide clues as to what might determine this point, but the question of when this happens remains unanswered.

Ferreira also asks for clarification of the distinction between one- and two-proposition sentences that figure prominently in our target article (see also **Martin**). In our work, a proposition expresses a set of thematic roles. (We are not stipulating that this is how to define the term, only indicating that this is what it refers to in our research to date.) We agree with Ferreira that whether a sentence such as *The boy hugged the girl and the baby* expresses one action or two depends on the mental model that the listener constructs. This is determined by the pictures in sentence–picture matching and is internally generated in object manipulation, plausibility judgment, and answering questions. Because the number-of-propositions effect arises at the conceptual or perceptual, not the linguistic, level, our prediction is that, if listeners are induced to form a mental model that involves two actions for a sentence such as *The boy hugged the girl and the baby,* by, say, adding *then* after *and,* then the sentence will pattern with two-proposition sentences.

We recognize that there is no full theory of the set of processes that constitute interpretive and post-interpretive processing and of their boundary. The distinction is admittedly vague enough to allow researchers to use it to escape from unwelcome results, but we think that it is also potentially rich and precise enough to serve a useful function in thinking about results. The same is true of the concept of working memory, as the discussion above demonstrates, and many other broad concepts.

R3. The relationship between working memory and interpretive and post-interpretive processes

We note that several commentators considered it plausible on theoretical grounds that there be more than one processing resource system utilized by verbal tasks (see **Wingfield** and **Toomela & Allik;** see also **Christiansen & McDonald**). We agree with this perspective, but will focus here on empirical issues.

R3.1. Sentence comprehension and working memory capacity in normal subjects. Many of the commentaries raised issues concerning the interpretation of the studies with normal subjects. The majority of these commentaries dealt with methodological and statistical issues that can affect the power of particular studies.

One question raised by several commentators concerned whether many of the studies cited have sufficient power to detect the predicted interactions. **Whitney & Budd** point out that it is difficult to test for interactions if interpretive processes are low in resource demands. In addition, they claim that designs that treat span as a grouping variable with three levels result in a substantial loss of power compared to those using only high or low spans or those treating span as a continuous variable. **Miyake et al.** make similar claims and argue that a major cause of low statistical power is the use of highly discrete measures of individual differences and the creation of arbitrary span groups. They point out that "discrete span measures have lower power because they reduce the variance by not capturing subtle differences that may exist among individuals with the same span score." They speculate that "in many cases (if not all), [our] failure to detect theoretically relevant interactions was due to insufficient power." In addition, they argue that "these nonsignificant trends likely will reach statistic significance if they are tested in multiple regression analyses, using a continuous measure of working memory span."

As we pointed out in the target article, we agree that researchers must be extremely careful when interpreting nonsignificant results (the "unenviable position" that Wingfield describes). However, several aspects of the data give us confidence in our interpretation. In the vast majority of cases we cite, the relevant interactions do not just fail to reach significance but in fact are far from significant. Furthermore, in our own studies (particularly of normal individuals) we often replicate the pattern of results over several studies using large samples (n = 100) and a new set of stimulus materials in each study. It is true that, if the same total number of subjects were tested but divided into two rather than three groups, the number of subjects per group would be larger, thus increasing power. However, the use of just two groups to make inferences about a continuous variable such as span can be highly problematic from a design standpoint. If only two groups are tested (high and low span) and there is a systematic relationship between span and some measure of language processing, the assumption is that this relationship is monotonic and that subjects who fall between high- and low-span groups in terms of working memory capacity would also fall between in terms of language-processing ability. However, we have found that in many instances subjects who fall between high- and low-span subjects in terms of working memory capacity perform either more poorly than low-span subjects or better than high-span subjects in terms of language-processing efficiency (see Figs. 1–3 in the target article). This finding dramatically alters the conclusion that would have been drawn if only two groups had been tested, and it justifies the inclusion of at least three groups.

Miyake et al. argue for the use of continuous measures of working memory span rather than the creation of arbitrary span groups. Although this approach is rarely taken by researchers in the field, we agree with it fully and have in fact used it in some of our own work. In our experience, the use of such measures does not increase the ability of traditional measures of working memory capacity to predict language-processing efficiency. In one study (Waters & Caplan 1996c), we tested 94 college students on a sentence span task four times, once with each of four different sentence types. All subjects were tested at all span sizes from 2 to 6 (i.e., 25 trials or 100 items). We investigated the relationship between measures of reading comprehension and various measures of span (total number of items correct, total number of trials correct, traditional working memory span). There was more variability in the item and trial measures, with the range of scores being 1–6 for the traditional span measure, but 1–100 for the item measure, and 1–25 for the trial measure. Standard deviations were approximately one-fourth of the mean for the trial and item measures. However, all three measures were highly correlated, and the greater range of scores did not result in significantly better correlations with the test of reading comprehension. Correlations between reading comprehension and the four traditional span measures ranged from .21 to .33; with the trial measure they ranged from .22 to .34 and with the item measure from .25 to .37. We have also taken this approach to the analysis of the on-line data from the auditory moving windows task presented in Figure 1 in the target article. In this analysis, rather than divide the subjects into discrete groups on the basis of working memory span, we examined the correlation between the number of items correctly recalled to span on the working memory task and the increase in processing time on object- compared to subject-relative sentences for each of the five phrases shown in Figure 1. All the correlations between this fine-grained measure of working memory capacity and this on-line measure of sentence-processing difficulty were nonsignificant (all correlations less than .10). An identical pattern of results was found in an additional study of 100 elderly subjects. Thus, whereas in principle the lack of relationship between traditional measures of working memory span and language processing could be due to the use of discrete groups based on sentence-final word recall, in practice the use of finer-grained measures of sentence-final word recall does not result in better predictive ability.

On the other hand, we have found that the ability of sentence span measures to predict language-processing efficiency does increase if not only sentence-final word recall but also efficiency (as indexed by reaction times (RTs) and errors) on the sentence-processing component of the sentence span task is taken into account when subjects are allocated to working memory groups. We have labeled this measure a composite Z score (Comp Z; Waters & Caplan 1996c). For example, in the study of college students outlined above, the correlation between the four different sentence span tasks and reading comprehension increased to .72 when sentence-processing efficiency was included in the calculation of working memory span. In a similar vein, differences between the high-, medium-, and low-span subjects shown in Figure 1 did emerge when the groups were divided on the basis of Comp Z.

A critical issue becomes how to interpret the better predictive power of a working memory measure that takes sentence-processing efficiency into account. **Kane et al.** argue that our "innovation" of including the sentence-processing component in the measure of working memory capacity pushes the measure toward being one of specific processing skills, as opposed to a measure of domain-free controlled attentional capacity. We think this is correct. Our data (Waters & Caplan 1996c) and those of Tirre and Pena (1992) suggest that most of the shared variance between sentence span tasks and reading comprehension tasks is accounted for by the processing component of the sentence span task. Daneman and Tardiff (1987) suggested that the entire relationship between the sentence span task and measures of language-processing efficiency can be accounted for by the overlap of operations in the two tasks (but see Waters & Caplan, 1996c, for some reservations about the basis for their conclusion). Our data (Waters & Caplan, 1996c) show that there also is a small but significant contribution of the recall measure to these correlations. For the reasons cited above, it is likely that the recall component of a working memory test is not a measure of a central verbal working memory capacity. The independent contribution of the recall measure in complex span tasks to predicting comprehension probably reflects a functional ability that overlaps between recall and certain types of memory requirements of text and discourse processing, such as those that are involved in finding the referents of pronouns and other anaphoric elements. We suggest that the major determinant of the relationship between sentence span working memory measures and text and discourse comprehension is the fact that the two involve structuring sentences and assigning their meanings, and we suggest that an additional contribution to the correlation is the fact that the two tasks share the need to refer to items held explicitly in a short-term memory system.

Kane et al. argue that the best way to measure the domain-free executive component of working memory is to administer a battery of different working memory tasks that share the dual-task quality but differ in domain-specific processes. They claim that domain-free working memory capacity is needed only under attention-demanding circumstances and that, because syntactic processing appears to be immune to divided attention, it likely occurs relatively automatically. **Kemper & Kemtes** have also taken the approach of using the data from several working memory tasks as a measure of working memory capacity. They claim that the data from such tasks suggest a single latent variable across age groups, but unlike Kane et al., they have found that such measures are related to interpretive processing, at least on the output side.

We have taken the approach suggested by **Kane et al.** in several studies. In one we tested 112 healthy elderly subjects (50–85 years old) on seven operation span measures (alphabet span, backward digit span, missing digit span, subtract 2, running item span, and two versions of sentence span). Correlations among all the measures (other than missing digit span) were moderate, even when the effects of age were partialled out. Correlations between the Nelson Denny reading comprehension subtest and the operation span scores ranged from −.04 to .30, the highest correlation being with the sentence span measure. However, there were no significant correlations between any of these measures and an index of processing difficulty on an on-line measure of auditory sentence processing (self-paced listening). These data support the notion that domain-free working memory plays very little role in on-line interpretive processing and provide further support for our claim, outlined above, that the relationship between sentence span measures and text and discourse comprehension likely has to do with the memory requirements of these tasks.

Walenski & Swinney raise a different set of methodological issues regarding the reading or sentence span task itself. They suggest that previously unexamined sources of variability may contribute to the lack of strong correlation between verbal working memory capacity and syntactic complexity. In particular, they argue that the materials used in the assessment of span should be controlled for their syntactic complexity and number of propositions. Furthermore, the commentators argue that the finding of better correlations for propositions than for complexity may be due to the fact that the span materials differ less in propositional content than in structural complexity. Although it is true that, in general, other researchers do not control the linguistic properties of the materials used in the sentence span task, these variables were either manipulated or controlled for in all of our studies with patients and normal individuals cited in the target article. In all these studies, subjects were divided into groups using a variant of the Daneman and Carpenter task that we devised in 1987 (Waters et al. 1987), in which the trials are blocked by sentence type and the sentences are systematically varied in terms of syntactic complexity and number of propositions. Thus, contrary to Walenski & Swinney's claim, these sources of variability were not unexamined and cannot account for the lack of strong correlation.

In addition to questions about the sensitivity of the working memory measures, some commentators questioned whether the chronometric and behavioral methods in the studies we cite are sensitive enough to detect the expected effects. **Kutas & King** claim that they are not. In particular, they claim that reaction time effects seen in the behavioral data occur after the point at which they would be expected (**Christiansen & MacDonald** seem to disagree). Kutas & King claim that event-related potential (ERP) data are more sensitive and in fact show the predicted complexity by group interactions. However, in the Kutas & King studies cited, the subject groups were based not on measures of working memory capacity but rather on comprehension scores. They claim that it is likely that good comprehenders would score better on working memory measures. However, the correlations found between comprehension and working memory measures are often at best only moderate (see data referred to above) and do not justify using comprehension scores in lieu of working memory scores if the question of interest is the relationship between working memory capacity and language processing.

Gibson & Roberts point out that, to date, the range of syntactic structures tested has been quite narrow. They argue that a metric, such as that developed in Gibson's model, can be used to make predictions about the relative difficulty of various syntactic structures. We agree and are in the process of extending our work to other structures. Gibson & Roberts also point out that they have found additional support for our hypothesis that working memory measures predict post-interpretive processing through an experiment in which subjects answered questions about sentences with varying numbers of propositions. **Lewis** develops a different model of the determinants of syntactic complexity; our impression is that the revision of Gibson's model found in his recent work accommodates the discrepant data Lewis pointed to in Gibson's earlier work. Lewis suggests that similarity of to-be-remembered items affects performance in all tasks that have a memory component. Perhaps this is true, but different types of similarity affect different memory performances differently (e.g., semantic similarity has little effect on immediate recall, and phonological similarity has a major effect; at longer delays, the effects are reversed), and these different effects of similarity considered along different dimensions are part of the basis for postulating different memory systems.

Walenski & Swinney point out that the evidence for a lack of correlation of verbal working memory with interpretive processes has come mainly from tasks that present materials visually and point out that the reading span task is also typically presented visually. They have found that sentence span scores are significantly higher when the test materials are presented auditorily and attribute this effect to the greater automaticity of the auditory modality. However, differences in sentence span scores in the auditory versus the visual modality are typically very small, and there is no evidence that subjects process the material any more automatically when the materials are presented auditorily. Walenski & Swinney also suggest that visually based language tasks may involve more post-interpretive processing. They make a plea for using only materials presented in the auditory modality at normal rates, suggesting that the auditory moving windows task may involve some task-induced post-perceptual processing, because the rate of presentation is slower than normal speech and so may not reflect automatic comprehension processes. However, the behavioral data do not suggest that the task picks up post-perceptual processing. Furthermore, the implication is that other auditory tasks might be a better reflection of automatic comprehension processes. However, all tasks other than simply listening to normal speech probably change the nature of the comprehension process. For example, in the cross-modal naming and lexical decision/priming tasks, the stimulus sentences are presented at a normal rate, but the presentation is interrupted by the presentation of a visual cue to which the subject must respond. This divided-attention task likely changes the manner in which subjects process the stimulus sentences.

Several commentaries claim that additional data not cited in the target article point to a relationship between working memory capacity and syntactic comprehension ability. **Pearlmutter** points out that Pearlmutter and MacDonald (1995) showed a span-by-ambiguity interaction in a self-paced reading task. High-span subjects were sensitive to plausibility in both unambiguous and ambiguous sentences, low-span subjects only in unambiguous sentences. Strategic factors induced by questions were not likely to have been present because, unlike the case in the study by MacDonald, Just and Carpenter (1992), the comprehension questions did not ask about the ambiguity. This study raises very interesting issues. We do not deny that there are differences in the verbal talents of high- and low-span subjects. Pearlmutter and MacDonald (1995) suggest that one such difference lies in their use of plausibility information in on-line sentence processing. Another possibility is that the difference lies in their judgments of probabilities of thematic roles, knowledge of lexical co-ocurrence statistics, and so on. It is possible that all subjects use information of this type on line; the difference between subjects might reside not in their spans but in their knowledge. To explore this, probability ratings would have to be obtained from the same subjects who are tested for on-line comprehension and the factors of individual probability ratings and working memory span examined for their independent relative contributions to the on-line sentence-processing measure.

Andrews & Halford claim to have found a complexity-by-capacity interaction when complexity is measured in a different way – one that takes the number of role assignments into account. Their account rests on the unsupported and counterintuitive view that subjects process all thematic roles in subject-object and doubly center-embedded object-relativized sentences at once but treat subject-relativized sentences with the same number of thematic roles on a clause-by-clause basis. In addition, the results described by Andrews & Halford are complex, with effects of working memory span appearing for different sentence types in inconsistent ways. Our reaction to these data is that they are hard to interpret and, although not supportive of our hypothesis, do not constitute results strongly contradictory to the data we reviewed in our target article.

Bates et al. cite Kilborn (1991), who found that subjects did not use morphological markings normally under conditions of noise. In our opinion, the experiment suffered significantly from the inclusion of ungrammatical sequences that the subjects had to interpret (e.g., *hits the wristwatch the television*), which predisposes subjects towards the adoption of problem-solving strategies, and from an inadequate number of examples of each type, which precludes the analysis of the data by items. These problems aside, the finding that morphological endings are not used normally when materials are presented in noise only suggests that they are not perceived in these conditions, which is hardly surprising given their lack of stress. Bates et al. also mention a paper by Blackwell and Bates (1995) as evidence that a concurrent digit load selectively disrupts recognition of subject–verb-agreement errors compared to omissions or transpositions of elements in a sentence. However, none of the interactions of error type and digit load was significant in the analyses.

Finally, **Miyake et al.** claim that the logic of the dual-task studies we cite is flawed because the digit span task interferes with the phonological loop component of working memory rather than imposing an external load on the central executive (CE). This is not the interpretation that other researchers, beginning with Baddeley and Hitch (1974), have made of the effect of a concurrent digit load task. One possible problem with the external load paradigm, which we acknowledge in the target article, is that it may not be demanding enough. We suggest that perhaps more demanding secondary tasks, such as random number generation, would prove more useful. However, **Towse et al.** point out that random number generation "is a product of complex and heterogeneous skills, difficult to render into a meaningful, singular construct." Given the dissociation over both normal and brain-damaged subjects of performance on tasks that are thought to measure CE functions (as **Gibson & Roberts** noted), we will have to rely on the slow accumulation of data using different interference paradigms, none of which is an ideal concurrent task to occupy the CE of a working memory system.

R3.2. Neuropsychological evidence. Most commentators felt that the data showing that patients with low working memory capacity had good syntactically based comprehension were among the strongest points in favor of our separate-sentence-interpretation-resource (SSIR) hypothesis. However, a few commentators disagreed, and several raised issues about other aspects of patients' performances.

Bates et al. point out that deficits in syntactic processing in patients with aphasia extend beyond difficulties with passives and object relatives to include processing morphological forms and that these difficulties are not restricted to patients with agrammatic speech. We agree with both these observations but believe they are not relevant to our thesis.

Christiansen & MacDonald argue that the data from patients with low working memory spans are mostly off-line accuracy measurements. They refer to results of theirs that indicate that patients with Alzheimer's disease (AD) are normal in cross-modal naming only when all but a few words in a sentence can be ignored and that their abilities to produce and interpret pronouns correlate well with working memory measures. Not having seen these data, we cannot discuss them in detail. We can say, however, that we have found normal effects of syntactic structure on self-paced listening in Parkinsonian patients with reduced working memory (studies with AD patients are underway).

Kolk & Hartsuiker discuss the possible use of "strategies" on the part of patients. They argue that a strategy that takes the first noun of a sentence as the agent of all the verbs in a sentence would account for the performance of aphasic subjects, particularly in object–subject-relative clauses (*The horse kicked the elephant that touched the dog*). We have long argued that aphasic performance is determined by a combination of a patient's deficit and the strategies that he or she uses (Caplan & Futter 1986; Caplan et al. 1985; Caplan & Hildebrandt 1987; 1988). In English, there are two well-documented strategies that aphasic patients use to assign thematic roles when syntactic analysis fails. Aphasic patients may either take the first noun in the sentence as the agent of every verb (the strategy Kolk & Hartsuiker mention) or take the noun immediately before each verb as the agent of that verb. The first strategy leads to errors in object–subject-relative sentences and good performance on conjoined sentences; the second leads to the opposite pattern. Both strategies are commonly used, and we have not been able to find a predilection for one or the other across large groups of aphasic patients. Individual patients tend to use one or the other, not both. Contrary to Kolk & Hartsuiker's claim about data of ours (Caplan & Waters 1996), accuracy on object–subject and conjoined sentences does not differ systematically in aphasic patients overall.

Kolk & Hartsuiker also discuss patients with AD and Parkinson's disease (PD). They claim that problems these patients have with sentences containing more than one proposition are related to a general verbal working memory limitation – our point in the target article exactly. They also say that these patients do not use word-order strategies and that this deficit is related to a "prefrontal dysfunction . . . in the domain of supervisory processes." To our knowledge, there is very little that can be said with confidence about whether AD and PD patients use the same types of strategies as aphasic patients. The most convincing data on strategies used by aphasics come from enactment tasks in which the patients can make whatever responses they choose; patients with AD and PD have been tested mostly on sentence–picture-matching tasks in which the foils are preselected and spontaneously generated strategies may not have a chance to emerge. However, given the ubiquity of the strategies used by aphasics (they are used by children as well), there is every reason to think that AD and PD patients use them also. We expect that Kolk & Hartsuiker's speculations in this regard will prove incorrect.

McCarthy & Warrington review their work on the sentence-comprehension deficits of patients with short-term memory (STM) impairments, arguing that these patients were able to understand sentences on-line but had difficulties when utterances had to be reanalyzed. We agree with McCarthy & Warrington's point that STM cases have not shown syntactic comprehension deficits that can be related to their STM limitations (see Waters et al., 1991, and Caplan & Waters, 1980). We are puzzled by their statement that we have not disentangled the various components of the working memory system. We distinguished between impairments in the rehearsal and phonological storage capacities that are found in the STM patients Warrington & McCarthy and other researchers described and disorders of the CE component of working memory (to use Baddeley's terminology). We did so because, as we noted in our target article, some researchers (e.g., Just & Carpenter 1992) have argued that patients with disorders of rehearsal and phonological storage are not adequate tests of the claim that there is a single working memory resource, whereas patients with limited "central executive" working memory functions are. It is for this reason that we deemphasized this literature in our target article in favor of discussion of patient groups with CE limitations. We appreciate McCarthy & Warrington's drawing attention to this literature.

McCarthy & Warrington promote single case studies over "large-scale surveys" in which they say "elegant dissociations and statistical artifacts are difficult to disentangle." We think one takes data wherever they come from. Statistical analyses of large data sets, such as those referred to by **Kemper & Kemtes,** tell us things that experimental exploration of single cases cannot. Single case studies may reflect exceptions to widely found patterns of performance and therefore may be misleading regarding the general properties of a cognitive system. All studies have weaknesses, of which limitations on inferences that can be drawn because of the population studied represent only one (and often not the most important). The evidence must be considered study by study.

Kotz & von Cramon describe a potentially important patient with a reduced working memory but normal STM performances. They report that H.G. performed normally on syntactic comprehension tasks, had no particular difficulty with sentences with more than one proposition, and did well on the Token Test. These performances indicate that H.G. was able to retain considerable sentential semantic content (several propositions, adjectives arbitrarily associated with nouns) in memory and use that content in enactment and other tasks. These are the types of post-interpretive functions that we have found often to be affected in patients with working memory limitations, and we think that Kotz & von Cramon are right that this case suggests that a working memory limitation as measured on complex span task does not necessarily lead to this type of problem.

Kotz & von Cramon also recorded ERPs in their patient and found that syntactic (word order) and semantic (selectional restriction) violations did not result in ERP components that are normally present, and that the P600 was delayed to fast visually presented sentences. They interpret these data as evidence for slowing of language processing owing to working memory limitations. This might be the correct interpretation of these data (if so, they speak against our hypothesis), but we think it is too early to tell. The fact that H.G. performed normally despite the absence of normal ERP components raises questions about the functional role of these components. According to our non-practitioner reading of the ERP literature, a delay of the P600 to 1,000 msec may not be of theoretical significance. It would be important to know that any delays in ERP components were selective, not a general result of the lesion in the patient. We also note that this is an example in which a single case cannot resolve an issue (see our discussion of **Warrington & McCarthy** above). The two findings could be unrelated disorders. One way to explore this possibility is to see whether this finding is replicated across patients, or whether the magnitude of a working memory deficit is correlated with delays in the P600.

Bánréti refers to work by Ullman et al. (1997) that reported good rule-based morphological processing in AD patients and poor processing of this sort in PD patients. Ullman et al. related this pattern to the corresponding integrity and impaired nature of procedural memory in these two groups of patients, and Bánréti endorses this hypothesis. Bánréti is right that Ullman et al.'s and our data regarding PD patients are not compatible if one attempts this synthesis. However, Ullman et al. studied word (past tense) formation, not sentential syntax. Whether the grand synthesis of all rule-based processing and procedural memory is valid is uncertain (cf. Dominey et al. 1997).

Martin, like **Kotz & von Cramon,** presents data from a single case that speak to the relationship between STM and sentence comprehension. Patient A.B. (Martin & Romani 1994) had problems with the retention of semantic information in immediate recall tasks and with comprehension of sentences in which semantic information had to be retained in memory for short periods before being associated with a construct. Martin argues that A.B.'s comprehension problem is part of interpretive processing and reflects a limitation in lexicosemantic mechanisms that support performance on STM tasks. Assuming that A.B.'s performance reflects a common co-occurrence of problems (as always, replication of a single case would be reassuring), it suggests that there are mechanisms common to retaining unstructured lexical semantic information in comprehension and in span (as Martin & Romani, 1994, conclude). One possibility is that there are severe limitations on the ability to retain unstructured lexical semantic information in comprehension. Gibson's model of the difficulties imposed by certain syntactic structures is consistent with this suggestion.

This raises the larger issue of the mechanisms that support span. Subjects activate and retain phonological and lexical representations in span. The report by **Toomela & Allik** of reduced span for the same noun in different cases in Estonian suggests that span is sensitive to morphological representations. It is tempting to argue that subjects analyze the stimuli in span tasks to the extent to which they can linguistically and use all the representations that they extract from the stimuli to support their performance. The extent to which STM performance is entirely the result of the application of normal language-processing mechanisms to the linguistically degenerate stimuli presented in span and other STM tasks remains unclear.

R4. Neural mechanisms

We suggested that deficits seen after lesions and functional neuroimaging using ERPs, magnetic resonance imaging (fMRI), and positron emission tomography (PET) suggest that working memory is a dorsolateral frontal function, whereas sentence comprehension is a perisylvian function. **Friston** argues that two functionally separate systems must have separate neural substrates and proposes an experiment in which one could see whether there are brain regions in which regional cerebral blood flow (rCBF) is influenced by the interaction of syntactic complexity and external memory load, such regions, should they exist, being evidence against our SLIR model. Stowe (1997) carried out such an experiment, and there were brain regions in which rCBF was related to the interaction of the variables of sentence complexity and external load. Unfortunately, the interaction was not easily interpretable: blood flow changed in similar ways in some regions in both the low-complexity/low-load and the high-complexity/high-load conditions.

Several commentators disagree with our suggestion that "working memory" tasks involve dorsolateral frontal cortex. **McCarthy & Warrington** suggest that the perisylvian region may be "a core neuroanatomical substrate for 'working memory' in those tasks that have high loading on executive problem-solving abilities." The evidence they cite in support of this is the finding that patients with inferior parietal lesions have STM deficits, especially affecting storage functions. **Martin** claims that activation work by Fiez et al. (1996), and their review of the literature, indicates that the regions activated by sentence processing and by working memory tasks overlap.

We agree with **McCarthy & Warrington** that there is evidence from neuropsychology that storage and rehearsal of phonological representations involve inferior parietal lobe and Broca's area. PET and fMRI data suggest a similar picture (Zatorre 1992; see Demonet, 1996, for review). Tasks that require more "computation" on stored information – even a simple comparison of one representation with another, as in the *n*-back or missing span task – have consistently activated the dorsolateral frontal lobe (see, e.g., Petrides et al. 1993). Most researchers attribute dorsolateral prefrontal (DLPF) activation to executive function in working memory tasks and perisylvian activation to storage and rehearsal. However, we agree with **Martin** that the picture is not as simple as this. For instance, the study by Fiez et al. (1996) quoted by Martin tested storage, not manipulation, of verbal material (recall of five items minus a fixation baseline) and found increased rCBF mainly in DLPF, not perisylvian cortex (the rCBF increase in Broca's area was of borderline significance; decreases in rCBF were seen in the middle temporal gyrus and the insula). One analysis in the Petrides et al. (1993) study (the subtraction of a counting baseline from the externally ordered digit task) showed increased rCBF in Brodmann's area 40 of the left hemisphere. While this might have been due to the storage requirements of this task, we cannot be sure of this interpretation. Martin is right that, at present, the picture is not as clear as we, and others, have suggested, although we add that the view articulated above remains a distinct possibility.

This brings us to a reconsideration of the import of localization findings. We agree with **Friston** that two functionally separate systems must have separate neural substrates and also with **Gibson & Roberts**'s view that such evidence is extremely powerful when it exists. However, we think the argument from localization to separate functional systems is unidirectional at present. Evidence for gross neuroanatomic separation constitutes strong support for functional specialization, but evidence for gross neuroanatomic overlap is still compatible with functional specialization. The reason for this is the scale factor; it is entirely possible that, at a finer level of description, two functional systems can utilize different elements such as neurotransmitters within a single brain region, such that rCBF increases related to the two functions are not distinguishable by available techniques.

R5. Concluding remarks

"Working memory" is a catchy term and an appealing concept. Many tasks, from digit span through inspecting mental images, seem to be capacity limited, and the idea that these limitations arise because a short-duration memory system can retain only so much information while it effects computations is a powerful metaphor. However, the idea of an undifferentiated central executive that supplies resources to all verbal tasks is much less appealing. Our target article marshalled evidence that this type of model does not account for individual differences, interference effects, and performance of brain damaged subjects in syntactic processing in sentence comprehension. The commentaries on our article suggest that there is no evidence that commonly used tests of working memory capacity measure a shared resource pool. We have suggested that "these considerations do not undermine the concept of working memory as a pool of resources that is used to temporarily store and to operate on activated representations *in a given task.* They direct the question to the delineation of tasks. A major distinction that we and others have made is between one or more resource pools that may be used for aspects of on-

https://doi.org/10.1017/S0140525X99321786 Published online by Cambridge University Press

line psycholinguistic processing and one or more resource pools that may be used for other verbally-mediated tasks. How fine a fractionation of resource systems will be needed to account for capacity limitations in language processing and other verbally mediated cognitive functions is not known at this point (Waters & Caplan 1996c).

References

Letters "a" and "r" appearing before authors' initials refer to target article and response, respectively.

Abelson, R. A. & Prentice, D. (1997) Contrast tests of interaction hypotheses. *Psychological Methods* 2:315–28. [AM]

Abney, S. P. (1989) A computational model of human parsing. *Journal of Psycholinguistic Research* 18:129–44. [aDC]

Allik, J. & Bachmann, T. (1983) How bad is icon? *Behavioral and Brain Sciences* 6:12–13. [AT]

Allik, J. & Help, T. (1985) Comments on S. T. Parker's "A social-technological model for the evolution of language." *Current Anthropology* 26:626. [AT]

Almor, A., Kempler, D., MacDonald, M. C., Andersen, E. S. & Tyler, L. K. T. (1998) Why do Alzheimer's patients have difficulty with pronouns? Working memory, semantics, and pronouns in Alzheimer's disease (submitted). [MHC]

Baddeley, A. D. (1966) Short-term memory for word sequences as a function of acoustic, semantic, and formal similarity. *Quarterly Journal of Experimental Psychology* 18:362–65. [RLL]

(1976) *The psychology of memory.* Basic Books. [aDC]

(1986) *Working memory.* Oxford University Press. [aDC, RAM]

(1990a) The development of the concept of working memory: Implications and contributions of neuropsychology. In: *Neuropsychological impairments of short-term memory,* ed. G. Vallar & T. Shallice. Cambridge University Press. [aDC]

(1990b) *Human memory: Theory and practice.* Allyn and Bacon. [AT]

(1993) Working memory and conscious awareness. In: *Theories of memory,* ed. A. F. Collins, S. E. Gathercole, M. A. Conway & P. E. Morris. Erlbaum. [aDC]

(1995) Working memory. In: *The cognitive neurosciences,* ed. M. S. Gazzaniga. MIT Press. [aDC, AT]

Baddeley, A. D., Bressi, S., Della Sala, S., Logie, R. & Spinnler, H. (1991) The decline of working memory in Alzheimer's disease. *Brain* 114:2521–42. [aDC]

Baddeley, A. D. & Hitch, G. J. (1974) Working memory. In: *The psychology of learning and motivation, vol. 8: Advances in research and theory,* ed. G. A. Bower. Academic Press. [arDC, MJK, RLL, JNT]

(1994) Developments in the concept of working memory. *Neuropsychology* 8:485–93. [aDC]

Baddeley, A. D., Logie, R., Bressi, S., Della Sala, S. D. & Spinnler, H. (1986) Dementia and working memory. *Quarterly Journal of Experimental Psychology* 38A:608–18. [aDC]

Bates, E., Devescovi, A., Dronkers, N., Pizzamiglio, L., Wulfeck, B., Hernandez, A., Juarez, L. & Marangolo, P. (1994) Grammatical deficits in patients without agrammatism. *Brain and Language* 47:400–402. [EB]

Baum, S. (1991) Sensitivity to syntactic violations across the age span: Evidence from a word monitoring task. *Journal of Clinical Linguistics and Phonetics* 5:317–28. [aDC]

Bayles, K. A. (1982) Language function in senile dementia. *Brain and Language* 16:265–80. [aDC]

Berndt, R., Mitchum, C. & Haendiges, A. (1996) Comprehension of reversible sentences in "agrammatism": A meta-analysis. *Cognition* 58:289–308. [aDC]

Berwick, R. C. & Weinberg, A. (1984) *The grammatical basis of linguistic performance: Language use and acquisition.* MIT Press. [aDC]

Bever, T. (1970) The cognitive basis for linguistic structures. In: *Cognition and the development of language,* ed. J. R. Hayes. Wiley. [RLL]

Blackwell, A. & Bates, E. (1995) Inducing agrammatic profiles in normals. *Journal of Cognitive Neuroscience* 7:228–57. [EB, rDC]

Blanken, G., Dittmar, J., Haas, J. C. & Wallesch, C. W. (1987) Spontaneous speech in senile dementia and aphasia: Implications for a neurolinguistic model of language production. *Cognition* 27:247–74. [aDC]

Boland, J. E. & Cutler, A. (1996) Interaction with autonomy: Multiple output models and the inadequacy of the great divide. *Cognition* 58:309–20. [NJP]

Brown, R. G. & Marsden, C. D. (1988) Internal versus external cues and the control of attention in Parkinson's disease. *Brain* 111:323–45. [aDC]

(1991) Dual task performance and processing resources in normal subjects and patients with Parkinson's disease. *Brain* 114:215–31. [aDC]

Buckner, R. & Petersen, S. (1996) What does neuroimaging tell us about the role of prefrontal cortex in memory retrieval? *Seminars in the Neurosciences* 8:47–55. [RCM]

Budd, D., Whitney, P. & Turley, K. J. (1995) Individual differences in working memory strategies for reading expository text. *Memory and Cognition* 23:735–48. [PW]

Caplan, D. (1983) A note on the 'word-order problem' in agrammatism. *Brain and Language* 20:155–65. [HHHJK]

(1985) Comments on Fodor's *The modularity of mind:* A neo-Cartesian alternative. *Behavioral and Brain Sciences* 8:6–7. [aDC]

(1987) *Neurolinguistics and linguistic aphasiology: An introduction.* Cambridge University Press. [aDC]

Caplan, D., Alpert, N. & Waters, G. S. (in press) Effects of syntactic structure and propositional number on patterns of regional cerebral blood flow. *Journal of Cognitive Neuroscience.* [aDC]

Caplan, D., Baker, C. & Dehaut, F. (1985) Syntactic determinants of sentence comprehension in aphasia. *Cognition* 21:117–75. [arDC]

Caplan, D. & Futter, C. (1986) Assignment of thematic roles to NPs by an agrammatic aphasic. *Brain and Language* 27:117–35. [rDC, HHHJK]

Caplan, D. & Hildebrandt, N. (1986) Language deficits and the theory of syntax: A reply to Grodzinsky. *Brain and Language* 27:168–77. [rDC]

(1988) *Disorders of syntactic comprehension.* MIT Press/Bradford Books. [rDC, HHHJK]

Caplan, D., Hildebrandt, N. & Makris, N. (1996) Location of lesions in stroke patients with deficits in syntactic processing in sentence comprehension. *Brain* 119:933–49. [aDC]

Caplan, D., Hildebrandt, N. & Waters, G. S. (1994) Interaction of verb selectional restrictions, noun animacy and syntactic form in sentence processing. *Language and Cognitive Processes* 9:549–85. [arDC]

(1997) Determinants of sentence comprehension in aphasic patients in sentence- picture matching tasks. *Journal of Speech Language and Learning Research* 40:542–55. [aDC]

Caplan, D. & Waters, G. (1990) Short-term memory and language comprehension: A critical review of the neuropsychological literature. In: *Neuropsychological impairments of short-term memory,* ed. G. Vallar & T. Shallice. Cambridge University Press. [arDC, MJK, AM]

(1995) Aphasic disturbances of syntactic comprehension and working memory capacity. *Cognitive Neuropsychology* 12:637–49. [aDC]

(1996) Syntactic processing in sentence comprehension under dual-task conditions in aphasic patients. *Language and Cognitive Processes* 11:525–51. [arDC, HHHJK]

Carpenter, P. A. & Just, M. A. (1988) The role of working memory in language comprehension. In: *Complex information processing: The impact of Herbert A. Simon,* ed. D. Klahr & K. Kotovsky. Erlbaum. [aDC]

Carpenter, P. A., Just, M. A. & Shell, P. (1990) What one intelligence test measures: A theoretical account of the processing in the Raven Progressive Matrices Test. *Psychological Review* 97:404–31. [AM]

Carpenter, P. A., Miyake, A. & Just, M. A. (1994) Working memory constraints in comprehension: Evidence from individual differences, aphasia, and aging. In: *Handbook of psycholinguistics,* ed. M. Gernsbacher. Academic Press. [AW]

Case, R. (1985) *Intellectual development: Birth to adulthood.* Academic Press. [JNT]

Case, R., Kurland, M. & Goldberg, J. (1982) Operational efficiency and the growth of short term memory span. *Journal of Experimental Child Psychology* 33:386–404. [JNT]

Cattell, R. B. (1963) Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology* 54:1–22. [MJK]

Chapman, R. & Miller, J. (1984) *SALT: Systematic analysis of language transcripts.* University of Wisconsin Press. [SK]

Cheung, H. & Kemper, S. (1992) Competing complexity metrics and adults' production of complex sentences. *Applied Psycholinguistics* 13:53–76. [SK]

Chomsky, N. (1957) *Syntactic structures.* Mouton. [aDC]

(1981) *Lectures on government and binding.* Foris. [aDC]

(1986) *Knowledge of language.* Praeger. [aDC]

(1993) *The minimalist program.* MIT Press. [aDC]

(1997) Language and mind: Current thoughts on ancient problems. *Pesquisa Linguistica* 3:4. [ZB]

Christensen, A. L. (1974) *Luria's neuropsychological investigation: Text.* Munksgaard. [aDC]

Christiansen, M. H. (1996) Recursive sentence structure in connectionist networks. Paper presented at the Ninth Annual CUNY Conference on Sentence Processing, CUNY, New York. [aDC]

Cohen, J. (1978) Partialed products *are* interactions; partialed powers *are* curve components. *Psychological Bulletin* 85:858–66. [AM]

(1983) The cost of dichotomization. *Applied Psychological Measurement* 7:249–53. [AM]

(1988) *Statistical power analysis for the behavioral sciences, 2nd edition*. Erlbaum. [AM]
Cohen, L. & Mehler, J. (1996) Click monitoring revisited: An on-line study of sentence comprehension. *Memory and Cognition* 24:94–102. [aDC]
Conrad, R. (1963) Acoustic confusions and memory span for words. *Nature* 197:1029–30. [RLL]
Conway, A. R. A. & Engle, R. W. (1996) Individual differences in working memory capacity: More evidence for a general capacity theory. *Memory* 4:577–90. [MJK]
Conway, A. R. A., Tuholski, S. W. & Engle, R. W. (1998) Individual differences in working memory: Insights from the counting span task. Paper presented at the Meeting of the Midwestern Psychological Association, Chicago, IL. [MJK]
Cowper, E. A. (1976) *Constraints on sentence complexity: A model of syntactic processing.* Ph. D. thesis, Brown University. [RLL]
Cronbach, L. J. & Snow, R. E. (1977) *Aptitudes and instructional models: A handbook for research on interactions.* Irvington. [PW]
Daneman, M. & Carpenter, P. (1980) Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior* 19:450–66. [aDC, MJK, JNT, MW]
Daneman, M. & Merikle, P. M. (1996) Working memory and language comprehension: A meta-analysis. *Psychological Bulletin and Review* 3:422–33. [aDC, AW]
Daneman, M. & Tardif, T. (1987) Working memory and reading skill re-examined. In: *Attention and performance XII: The psychology of reading,* ed. M. Coltheart. Erlbaum. [rDC]
Davis, C. A. & Ball, H. E. (1989) Effects of age on comprehension of complex sentences in adulthood. *Journal of Speech and Hearing Research* 32:143–50. [aDC]
Della Sala, S. & Logie, R. H. (1993) When working memory does not work: The role of working memory in neuropsychology. In: *Handbook of neuropsychology, vol. 8,* ed. F. Boller & J. Grafman. Elsevier. [RAM]
Demonet, J. F., Fiez, J. A., Paulesu, E., Petersen, S. E. & Zatorre, R. J. (1996) PET studies of phonological processing: A critical reply to Poeppel. *Brain and Language* 55:352–79. [rDC]
Dennett, D. C. (1991) *Consciousness explained.* Little, Brown and Co. [AT]
DeRenzi, E. & Vignolo, L. (1962) The token test: A sensitive test to detect receptive disturbances in aphasics. *Brain* 85:665–78. [aDC, SAK]
Dick, F., Bates, E., Wulfeck, B. & Dronkers, N. (1998) *Deficits in the interpretation of complex sentences in aphasic patients and normals under adverse processing conditions.* University of California, San Diego, Center for Research in Language. [EB]
Dominey, P. F., Venre-Dominey, J., Brousolle, E. & Jeannerod, M. (1997) Analogical transfer is effective in a serial reaction time task in Parkinson's disease: Evidence for a dissociable sequence learning mechanism. *Neuropsychologia* 35:1–9. [rDC]
Emery, O. B. (1985) Language and aging. *Experimental Aging Research (Monograph)* 11:3–60. [aDC]
(1986) Linguistic decrement in normal aging. *Language and Communication* 6:47–64. [aDC]
Engle, R. (1995) Overview of working memory. Paper presented at the Conference on Cognitive Aging, Atlanta. [aDC]
Engle, R. W., Cantor, J. & Carullo, J. J. (1992) Individual differences in working memory and comprehension: A test of four hypotheses. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 18:972–92. [PW]
Engle, R. W. & Conway, A. R. A. (1998) Working memory and comprehension. In: *Working memory and thinking,* ed. R. Logie & K. Gilhooly. Psychology Press. [MJK]
Engle, R. W., Kane, M. J. & Tuholski, S. W. (in press a) Individual differences in working memory capacity and what they tell us about controlled attention, general fluid intelligence and functions of the prefrontal cortex. In: *Models of working memory: Mechanisms of active maintenance and executive control,* ed. A. Miyake & P. Shah. Cambridge University Press. [MJK]
Engle, R. W., Tuholski, S. W., Laughlin, J. E. & Conway, A. R. A. (in press b) Working memory, short-term memory and general fluid intelligence: A latent variable approach. *Journal of Experimental Psychology: General.* [MJK]
Eslinger, P. J. & Damasio, A. R. (1985) Severe disturbance of higher cognition after bilateral frontal lobe ablation: Patient EVR. *Neurology* 35:1731–41. [AM]
Farah, M. J. (1994) Neuropsychological inference with an interactive brain: A critique of the "locality" assumption. *Behavioral and Brain Sciences* 17:43–61. [PW]
Feier, C. D. & Gerstman, L. J. (1980) Sentence comprehension abilities throughout the adult life span. *Journal of Gerontology* 35:722–28. [aDC]
Ferreira, F. (1992) Linguistic processing is modular. Talk presented at the University of Wisconsin-Milwaukee Conference, Milwaukee, Wisconsin. [aDC]
Ferreira, F., Anes, M. D. & Horine, M. D. (1996) Exploring the use of prosody during language comprehension using the auditory moving window technique. *Journal of Psycholinguistic Research* 25:273–90. [aDC]
Ferreira, F., Henderson, J. M., Anes, M. D., Weeks, P. A., Jr. & McFarlane, D. K. (1996) Effects of lexical frequency and syntactic complexity in spoken language comprehension: Evidence from the auditory moving window technique. *Journal of Experimental Psychology: Learning, Memory and Cognition* 22:324–35. [aDC]
Fiez, J., Raife, E., Balota, D., Scheartz, J., Raichle, M. & Peterson, S. (1996) A positron emission tomography study of the short-term maintenance of verbal information. *Journal of Neuroscience* 16:808–22. [rDC, RCM]
Fodor, J. A. (1983) *The modularity of mind.* MIT Press. [arDC, NJP, PW]
Fodor, J. A., Bever, T. G. & Garrett, M. (1974) *The psychology of language: An introduction to psycholinguistics and generative grammar.* McGraw Hill. [MW]
Fodor, J. D. & Inoue, A. (in press) Attach anyway. In: *Reanalysis in sentence processing,* ed. F. Ferreira & J. D. Fodor. Kluwer. [FF]
Ford, M. (1983) A method for obtaining measures of local parsing complexity throughout sentences. *Journal of Verbal Learning and Verbal Behavior* 22:203–18. [aDC]
Forster, K. I. (1979) Levels of processing and the structure of the language processor. In: *Sentence processing,* ed. W. E. Cooper & E. C. T. Walker. Erlbaum. [aDC]
Frauenfelder, U., Segui, J. & Mehler, J. (1980) Monitoring around the relative clause. *Journal of Verbal Learning and Verbal Behavior* 19:328–37. [aDC]
Frazier, L. (1985) Syntactic complexity. In: *Natural language parsing: Psychological, computational, and theoretical perspectives,* ed. D. R. Dowty, L. Karttunen & A. M. Zwicky. Cambridge University Press. [SK]
(1990) Exploring the architecture of the language-processing system. In: *Cognitive models of speech processing: Psycholinguistic and computational perspectives,* ed. G. T. M. Altmann. MIT Press/Bradford Books. [aDC]
Frazier, L. & Clifton, C. (1996) *Construal.* MIT Press. [arDC]
Frazier, L. & Rayner, K. (1982) Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology* 143:178–210. [aDC]
Freidrich, F. J. (1990) Multiple phonological representations and verbal short-term memory. In: *The neuropsychology of short-term memory,* ed. T. Shallice & G. Vallar. Cambridge University Press. [aDC]
Friederici, A. D., Hahne, A. & von Cramon, D. Y. (in press) First-pass versus second-pass parsing processes in a Wernicke's and a Broca's aphasic: Electrophysiological evidence for a double dissociation. *Brain and Language.* [SAK]
Frick, R. W. (1995) Accepting the null hypothesis. *Memory and Cognition* 23:132–38. [AM]
Friston, K. J., Price, C. J., Fletcher, P., Moore, C., Frackowiak, R. S. J. & Dolan, R. J. (1996) The trouble with cognitive subtraction. *NeuroImage* 4:97–104. [KJF]
Gathercole, S. E. & Baddeley, A. D. (1993) *Working memory and language.* Erlbaum. [aDC, AM]
Gibson, E. (1991) *A computational theory of human linguistic processing: Memory limitations and processing breakdown.* Ph. D. thesis, Carnegie Mellon University, CMU-CMT-91–125. [RLL]
(in press) Linguistic complexity: Locality of syntactic dependencies. *Cognition.* [aDC, EG]
Gibson, E., Schutze, C. & Salomon, C. (1996) The relationship between the frequency and the processing of complex linguistic structures. *Journal of Psycholinguistic Research* 25:59–92. [SK]
Gibson, E. & Thomas, J. (1996) The processing complexity of English center-embedded and self-embedded structures. In: *Proceedings of the NELS 26 workshop on language processing: MIT Working Papers in Linguistics,* ed. C. Schutze. MIT Press. [aDC]
Gick, M. L., Craik, F. I. M. & Morris, R. G. (1988) Task complexity and age differences in working memory. *Memory and Cognition* 16:353–61. [aDC]
Grafton, S. T., Hazeltine, E. & Ivry, R. (1995) Functional mapping of sequence learning in normal humans. *Journal of Cognitive Neuroscience* 7:497–510. [aDC]
Grafton, S. T., Mazziota, J. C., Presty, S., Friston, K., Frackowiak, R. S. J. & Phelps, M. E. (1992) Functional anatomy of human procedural learning determined with regional cerebral blood flow and PET. *Journal of Neuroscience* 12:2542–48. [aDC]
Grober, E. & Bang, S. (1995) Sentence comprehension in Alzheimer's disease. *Developmental Neuropsychology* 11:95–107. [aDC]
Grodzinsky, Y. (1990) *Theoretical perspectives on language deficits.* MIT Press. [HHHJK]
Grossman, M., Carvell, S., Gollomp, S., Stern, M. B., Vernon, G. & Hurtig, H. I. (1991) Sentence comprehension and praxis deficits in Parkinson's disease. *Neurology* 41:1620–28. [aDC]
Grossman, M., Carvell, S., Stern, M. B., Gollomp, S. & Hurtig, H. (1992) Sentence

comprehension in Parkinson's disease: The role of attention and memory. *Brain and Language* 42:347–84. [aDC]
Halford, G. S., Wilson, W. H. & Phillips, S. (1998) Processing capacity defined by relational complexity: Implications for comparative, developmental, and cognitive psychology. *Behavioral and Brain Sciences* 21(6):803–65. [GA]
Hart, S. (1988) Language and dementia: A review. *Psychological Medicine* 18:99–112. [aDC]
Hawkins, J. A. (1994) *A performance theory of order and constituency.* Cambridge University Press. [RLL]
Hier, D. B., Hagenlocker, K. & Schindler, A. G. (1985) Language disintegration in dementia: Effects of etiology and severity. *Brain and Language* 25:117–33. [aDC]
Hildebrandt, N., Caplan, D. & Evans, K. (1987) The man "left" without a trace: A case study of aphasic processing of empty categories. *Cognitive Neuropsychology* 4:257–302. [aDC]
Hitch, G. J. & Baddeley, A. D. (1976) Verbal reasoning and working memory. *Quarterly Journal of Experimental Psychology* 28:603–21. [JNT]
Horn, J. L. (1980) Concepts of intellect in relation to learning and adult development. *Intelligence* 4:285–317. [MJK]
Horn, J. L. & McArdle, J. J. (1992) A practical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research* 18:117–44. [SK]
Humphreys, L. G. & Fleishman, A. (1974) Pseudo-orthogonal and other analysis of variance designs involving individual-differences variable. *Journal of Educational Psychology* 66:464–72. [AM]
Illes, J. (1989) Neurolinguistic features of spontaneous language production dissociate three forms of neurodegenerative disease: Alzheimer's, Huntington's, and Parkinson's. *Brain and Language* 37:628–42. [aDC]
Irigaray, L. (1973) *Le langage des dements.* Mouton. [aDC]
Irle, E. (1987) Lesion size and recovery of function: Some new perspectives. *Brain Research Reviews* 12:307–20. [AT]
(1990) An analysis of the correlation of lesion size, localization and behavioral effects in 283 published studies of cortical and subcortical lesions in old-world monkeys. *Brain Research Reviews* 15:11–213. [AT]
Johnson, M. (1996) Resource sensitivity in grammar and processing. Paper presented at the Ninth Annual CUNY Conference on Human Sentence Processing, CUNY, New York. [aDC]
Jonides, J., Smith, E. E., Awh, E., Minoshima, S. & Monton, M. A. (1993) Spatial working memory in humans as revealed by PET. *Nature* 363:623–25. [aDC]
Just, M. A. & Carpenter, P. A. (1992) A capacity theory of comprehension: Individual differences in working memory. *Psychological Review* 99(1):122–49. [GA, arDC, EG, MJK, HHHJK, AM, MW, PW]
Just, M. A., Carpenter, P. A. & Keller, T. (1996) Working memory: New frontiers of evidence. *Psychological Review* 103:773–75. [aDC]
Just, M. A. Carpenter, P. A., Keller, T. A., Eddy, W. F. & Thulborn, K. R. (1996b) Brain activation modulated by sentence comprehension. *Science* 274:114–16. [aDC]
Kahneman, D. (1973) *Attention and effort.* Prentice-Hall. [AW]
Kane, M. J. & Engle, R. W. (1988) Full frontal fluidity: Working memory capacity, attention, general fluid abilities, and the prefrontal cortex (submitted). [MJK]
Kapur, N. (1996) Paradoxical functional facilitation in brain-behaviour research. A critical review. *Brain* 119:1775–90. [AT]
Kemper, S. (1986) Imitation of complex syntactic constructions by elderly adults. *Applied Psycholinguistics* 7:277–87. [aDC]
(1987) Life-span changes in syntactic complexity. *Journal of Gerontology* 42:232-38. [Sk]
(1988) Geriatric psycholinguistics: Syntactic limitations of oral and written language. In: *Language, memory, and aging,* ed. L. L. Light & D. M. Burke. Cambridge University Press. [aDC, SK]
Kemper, S., Anagnopoulos, C., Lyons, K. & Heberlein, W. (1994) Speech accomodations to dementia. *Journal of Gerontology* 49:223–29. [aDC]
Kemper, S., Kynette, D. & Norman, S. (1992) Age differences in spoken language. In: *Everyday memory and aging,* ed. R. West & J. Sinnott. Springer-Verlag. [SK]
Kemper, S., Kynette, D., Rash, S., O'Brien, K. & Sprott, R. (1989) Life-span changes to adults' language: Effects of memory and genre. *Applied Psycholinguistics* 10:49–66. [SK]
Kemper, S., LaBarge, E., Ferraro, F. R., Cheung, H., Cheung, H. & Storandt, M. (1993) On the preservation of syntax in Alzheimer's disease: Evidence from written sentences. *Archives of Neurology* 50:81–86. [aDC]
Kemper, S. & Rash, S. (1988) *Practical aspects of memory: Current research and issues.* Cambridge University Press. [aDC]
Kemper, S., Snowdon, D. A. & Greiner, L. H. (1996) Linguistic ability across the life span: Findings from the Nun study. Paper presented at the Cognitive Aging Conference, Atlanta. [aDC]
Kempler, D., Curtiss, S. & Jackson, C. (1987) Syntactic preservation in Alzheimer's disease. *Journal of Speech and Hearing Research* 30:343–50. [aDC]
Kemtes, K. A. & Kemper, S. (1996) Younger and older adults' working memory and on- line processing of syntactically ambiguous sentences. Paper presented at the Cognitive Aging Conference, Atlanta. [aDC]
(1998) Age-based construct validation of working memory. Unpublished manuscript. [SK]
Kilborn, K. (1991) Selective impairment of grammatical morphology due to induced stress in normal listeners: Implications for aphasia. *Brain and Language* 41:275–88. [EB, rDC]
King, J. & Just, M. A. (1991) Individual differences in syntactic processing: The role of working memory. *Journal of Memory and Language* 30:580–602. [aDC, PW]
King, J. W. & Kutas, M. (1995) Who did what and when: Using word- and clause-level ERPs to monitor working memory usage in reading. *Journal of Cognitive Neuroscience* 7(3):376–95. [MK]
Kintsch, W. & Keenan, J. M. (1973) Reading rate and retention as a function of the number of the propositions in the base structure of sentences. *Cognitive Psychology* 5:257–74. [SK]
Kirshner, H. S. (1982) Language disorders in dementia. In: *The neurology of aphasia,* ed. H. S. Kirshner & F. R. Freeman. Swets and Zeitlinger. [aDC]
Kluender, R. & Kutas, M. (1993) Subjacency as a processing phenomenon. *Language and Cognitive Processes* 8:573–633. [aDC]
Kolk, H. H. J. & van Grunsven, M. F. (1985) Agrammatism as a variable phenomenon. *Cognitive Neuropsychology* 2:347–84. [HHHJK]
Kolk, H. H. J. & Weijts, M. (1996) Judgements of semantic anomaly in agrammatic patients: Argument movement, syntactic complexity and the use of heuristics. *Brain and Language* 54:86–135. [HHHJK]
Kontiola, P., Laaksonen, R., Sulkava, R. & Erkinjuntii, T. (1990) Pattern of language impairment is different in Alzheimer's disease and multi-infarct dementia. *Brain and Language* 38:364–83. [aDC]
Koster, J. (1975) Dutch as an SOV language. *Linguistic Analysis* 1:111–36. [HHHJK]
Krueger, L. E. (1994) Mind your *p*'s and *q*'s when accepting or rejecting the null hypothesis. Paper presented at the Psychonomic Society, St. Louis, MO. [aDC]
Kuno, S. (1974) The position of relative clauses and conjunctions. *Linguistic Inquiry* 5:117–36. [RLL]
Kutas, M. & King, J. W. (1996) The potentials for basic sentence processing: Differentiating integrative processes. In: *Attention and performance XVI,* ed. I. Ikeda & J. L. McClelland. MIT Press. [MK]
Kynette, D. & Kemper, S. (1986) Aging and the loss of grammatical forms. *Language and Communication* 6:65–72. [SK]
Lalami, L., Marblestone, K., Schuster, S., Andersen, E., Kempler, D., Tyler, L. & MacDonald, M. (1996) On-line vs. off-line sentence processing in Alzheimer's disease. Poster presented at the Meeting of Cognitive Aging, Atlanta. [aDC]
Lee, L. (1974) *Developmental sentence analysis.* Northwestern University Press. [SK]
Lees, A. J. & Smith, E. (1983) Cognitive deficits in the early stages of Parkinson's disease. *Brain* 106:257–70. [aDC]
Lehto, J. (1996) Are executive function tests dependent on working memory capacity? *Quarterly Journal of Experimental Psychology* 49A:29–50. [AM]
Lewis, R. L. (1993) *An architecturally-based theory of human sentence comprehension.* Ph. D. thesis, School of Computer Science, Carnegie Mellon University (CMU-CS- 93–226). [RLL]
(1996a) A theory of grammatical but unacceptable embeddings.Unpublished. [aDC]
Lewis, R. L. (1996b) Interference in short-term memory: The magical number two (or three) in sentence processing. *Journal of Psycholinguistic Research* 25:93–115. [SK, RLL]
(1998) Working memory in sentence processing: Retroactive and proactive interference in parsing. Talk presented at CUNY 98, New Brunswick, NJ. [RLL]
Lieberman, P., Friedman, J. & Feldman, L. S. (1990) Syntax comprehension in Parkinson's disease. *Journal of Nervous and Mental Diseases* 178:360–66. [aDC]
Light, L. L. (1990) Interactions between memory and language in old age. In: *Handbook of the psychology of aging, 3rd edition,* ed. J. E. Birren & K. W. Schale. Academic Press. [aDC, AW]
Logie, R., Zucco, G. M. & Baddeley, A. D. (1990) Interference with visual short-term memory. *Acta Psychologica* 75:55–74. [RLL]
Luria, A. R. (1974) *Neiropskhologiya pamyati. Narusheniya pamyati pri lokal 'nykh porazheniakh mozga.* [Neuropsychology of memory. Memory disturbances in local brain lesions]. Pedagogika. [AT]
MacDonald, M. C. (1997) *Language and Cognitive Processes: Special Issue on Lexical Representations and Sentence Processing* 12:121–399. [aDC]
MacDonald, M. C. & Christiansen, M. H. (1998) Individual differences without working memory: A reply to Just & Carpenter and Waters & Caplan. (submitted). [MHC, NJP]

MacDonald, M. C., Just, M. A. & Carpenter, P. A. (1992) Working memory contraints on the processing of syntactic ambiguity. *Cognitive Psychology* 24:56–98. [arDC]

MacDonald, M. C., Pearlmutter, N. J. & Seidenberg, M. S. (1994) The lexical nature of syntactic ambiguity resolution. *Psychological Review* 101:676–703. [aDC, AM]

Magnussen, S., Greenlee, M. W., Asplund, R. & Dyrnes, S. (1991) Stimulus-specific mechanisms of visual short-term memory. *Vision Research* 31:1213–19. [RLL]

Marslen-Wilson, W. D. (1987) Functional parallelism in spoken word-recognition. *Cognition* 25:71–102. [aDC]

Marslen-Wilson, W. D. & Tyler, L. K. (1987) Against modularity. In: *Modularity in knowledge representation and natural language understanding,* ed. J. L. Garfield. MIT Press. [NJP]

Martin, R., R. (1995) Working memory doesn't work: A critique of Miyake et al.'s capacity theory of aphasic comprehension deficits. *Cognitive Neuropsychology* 12:623–36. [RCM]

Martin, R. & Feher, E. (1990) The consequences of reduced memory span for comprehension of semantic versus syntactic information. *Brain and Language* 38:1–20. [aDC, SAK]

Martin, R. C. & Romani, C. (1994) Verbal working memory and sentence processing: A multiple component view. *Neuropsychology* 8:506–23. [rDC, RCM]

Martin, R. C., Shelton, J. & Yaffee, L. (1994) Language processing and working memory: Neuropsychological evidence for separate phonological and semantic capacities. *Journal of Memory and Language* 33:83–111. [RCM]

Mattys, S. L. (1997) The use of time during lexical processing and segmentation: A review. *Psychonomic Bulletin and Review* 4:310–29. [AW]

Maxwell, S. E. & Delaney, H. D. (1993) Bivariate median splits and spurious statistical significance. *Psychological Bulletin* 113:181–90. [AM]

Maxwell, S. E., Delaney, H. D. & Dill, C. A. (1984) Another look at ANCOVA versus blocking. *Psychological Bulletin* 95:136–47. [AM]

McCarthy, R. A. & Warrington, E. K. (1987a) The double dissociation of short-term memory for lists and sentences. *Brain* 110:1545–64. [RAM]

(1987b) Understanding: A function of short-term memory? *Brain* 110:1565–78. [RAM]

(1990a) *Cognitive neuropsychology: A clinical introduction.* Academic Press. [RAM]

(1990b) Short-term memory: A phenomenon in search of a function? In: *Neuropsychological impairments of short-term memory,* ed. G. Vallar & T. Shallice. Cambridge University Press. [RAM]

McClelland, G. H. (1997) Optimal design in psychological research. *Psychological Methods* 2:3–19. [AM]

McClelland, G. H. & Judd, C. M. (1993) Statistical difficulties of detecting interactions and moderator effects. *Psychological Bulletin* 114:376–90. [AM]

Mecklinger, A., Schriefers, H., Steinhauer, K. & Friederici, A. D. (1995) Processing relative clauses varying on syntactic and semantic dimensions: An analysis with event-related potentials. *Memory and Cognition* 23(4):477–94. [MK]

Miller, G. A. & Chomsky, N. (1963) Finitary models of language users. In: *Handbook of mathematical psychology, vol. II,* ed. R. R. Bush & E. Galanter. Wiley. [RLL]

Miyake, A., Carpenter, P. & Just, M. (1994) A capacity approach to syntactic comprehension disorders: Making normal adults perform like aphasic patients. *Cognitive Neuropsychology* 11:671–717. [EB, aDC, AM]

Morris, R. G. (1984) Dementia and the functioning of the articulatory loop system. *Cognitive Neuropsychology* 1:143–57. [aDC]

(1986) Short-term forgetting in senile dementia of the Alzheimer's type. *Cognitive Neuropsychology* 3:77–97. [aDC]

(1987) Articulatory rehearsal in Alzheimer type dementia. *Brain and Language* 30:351–62. [aDC]

Morris, R. G. & Baddeley, A. D. (1988) Primary and working memory functioning in Alzheimer-type dementia. *Journal of Clinical and Experimental Neuropsychology* 10:279–96. [aDC]

Morris, R. G., Craik, F. I. M. & Gick, M. L. (1990) Age differences in working memory tasks: The role of secondary memory and the central executive system. *Quarterly Journal of Experimental Psychology* 42A:67–86. [aDC]

Morris, R. G., Gick, M. L. & Craik, F. I. M. (1988) Processing resources and age differences in working memory. *Memory and Cognition* 16:362–66. [aDC]

Müller, H. M., King, J. W. & Kutas, M. (1997) Event-related potentials elicited by spoken relative clauses. *Cognitive Brain Research* 5(3):193–203. [MK]

Münte, T. F., Schwirtz, O., Wieringa, B. M., Matzke, M. & Johannes, S. (1997) Electrophysiology of complex sentences: Word and sentence-level ERPs. *EEG-EMG-Zeitschrift für Elektroenzephalographie, Elektromyographie und verwandte Gebiete* 28(1):11–17. [MK]

Natsopoulos, D., Katsarou, Z., Bostantzopoulo, S., Grouios, G., Mentanopoulos, G. & Logothetis, J. (1991) Strategies in comprehension of relative clauses in Parkinsonian patients. *Cortex* 27:255–68. [aDC]

Naucler, N., Wulfeck, B. & Bates, E. (1998) *A developmental study of complex sentence interpretation abilities* (Technical Report CND-9803). University of California, San Diego, Project in Cognitive and Neural Development. [EB]

Navon, D. (1984) Resources – A theoretical soup stone. *Psychological Review* 91:216–34. [rDC, MHC]

Neville, H., Nicol, J. L., Barss, A., Forster, K. I. & Garrett, M. F. (1991) Syntactically based sentence processing classes: Evidence from event-related brain potentials. *Journal of Cognitive Neuroscience* 3:151–65. [aDC]

Nõmm, M. (1996) The development of short-memory span: Further evidence on the contribution of long-term memory. Unpublished Master's thesis, Tartu University. [AT]

Norman, S., Kemper, S., Kynette, D., Cheung, H. & Anagnopoulous, C. (1991) Syntactic complexity and adults running memory span. *Journal of Gerontology: Psychological Sciences* 46:346–51. [aDC]

Obler, L. K., Fein, D., Nicholas, M. & Albert, M. L. (1991) Auditory comprehension and aging: Decline in syntactic processing. *Applied Psycholinguistics* 12:433–52. [aDC]

Pearlmutter, N. J. & MacDonald, M. C. (1992) Plausibility and syntactic ambiguity resolution. In: *Proceedings of the 14th Annual Conference of the Cognitive Science Society.* Erlbaum. [NJP]

(1995) Individual differences and probabilistic constraints in syntactic ambiguity resolution. *Journal of Memory and Language* 34:521–42. [arDC, NJP]

Petrides, M., Alivisatos, B., Meyer, E. & Evans, A. (1993) Functional activation of the human frontal cortex during the performance of verbal working memory tasks. *Neurobiology* 90:878–82. [arDC]

Phillips, C. B., Zeki, S. & Barlow, H. B. (1984) Localization of function in the cerebral cortex. Past, present and future. *Brain* 107:327–61. [KJF]

Plaut, D., McClelland, J., Seidenberg, M. & Patterson, K. (1996) Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review* 103:56–115. [rDC]

Poizner, H., Bellugi, U. & Tweney, R. D. (1981) Processing of formational, semantic, and iconic information in American sign language. *Journal of Experimental Psychology: Human Perception and Performance* 7:1146–59. [RLL]

Posner, M. I., Peterson, S. E., Fox, P. T. & Raichle, M. E. (1988) Localization of cognitive operations in the human brain. *Science* 240:1627–31. [aDC]

Potter, M. C. (1993) Very short-term conceptual memory. *Memory and Cognition* 21:156–61. [AW]

Price, C. J. & Friston, K. J. (1997) Cognitive conjunction: A new approach to brain activation experiments. *NeuroImage* 5:261–70. [KJF]

Rakic, P. (1988) Specification of cerebral cortical areas. *Science* 241:170–76. [aDC]

Raz, N., Gunning-Dixon, F. M., Head, D., Dupois, J. H. & Acker, J. D. (1998) Neuroanatomical correlates of cognitive aging: Evidence from structural MRI. *Neuropsychology* 12:95–106. [AW]

Restle, F. (1970) Theory of serial pattern learning: Structural trees. *Psychological Review* 77:481–95. [AT]

Roberts, R. M. & Corkin, S. (1998) What do working memory tasks measure? Manuscript under review, Department of Brain and Cognitive Sciences, MIT. [EG]

Roberts, R. M., Hooker, C. & Gibson, E. (1997) Sentence memory capacity and working memory. Poster presented at the 10th CUNY Sentence Processing Conference, Santa Monica, CA, March 1997. [EG]

Rochon, E. & Saffran, E. M. (1995) A semantic contribution to sentence comprehension impairments in Alzheimer's disease. Poster presented to the Academy of Aphasia, San Diego. [aDC]

Rochon, E., Waters, G. S. & Caplan, D. (1994) Sentence comprehension in patients with Alzheimer's disease. *Brain and Language* 46:329–49. [aDC]

Rosenberg, S. & Abbeduto, L. (1987) Indicators of linguistic competence in the peer group conversational behavior of mildly retarded adults. *Applied Psycholinguistics* 8:19–32. [SK]

Rosenthal, R. R. & Rosnow, R. L. (1985) *Contrast analysis: Focused comparisons in the analysis of variance.* Cambridge University Press. [AM]

Salthouse, T. A. (1991) *Theoretical perspectives on cognitive aging.* Erlbaum. [aDC]

Schneider, W. (1993) Varieties of working memory as seen in biology and in connectionist/control architectures. *Memory and Cognition* 21:184–92. [AT]

Schneider, W. & Shiffrin, R. M. (1977) Controlled and automatic human information processing I.: Detection, search, and attention. *Psychological Review* 84:1–66. [aDC]

Schwartz, M. F., Marin, O. S. M. & Saffran, E. M. (1979) Dissociations of language function in dementia: A case study. *Brain and Language* 7:277–306. [aDC]

Schwartz, M. F., Saffran, E. M. & Marin, O. S. M. (1980) The word order problem in agrammatism I: Comprehension. *Brain and Language* 10:249–67. [RAM]

Seidenberg, M. S. & McClelland, J. L. (1989) A distributed, developmental model of word recognition and naming. *Psychological Review* 96:523–68. [rDC]

Shah, P. & Miyake, A. (1996) The separability of working memory resources for spatial thinking and language processing: An individual differences approach. *Journal of Experimental Psychology: General* 125:4–27. [aDC]

Shallice, T. & Burgess, P. (1993) Supervisory control of action and thought selection. In: *Attention: Selection, awareness, and control,* ed. A. Baddeley & L. Weiscrantz. Clarendon Press. [MJK]

(1996) The domain of supervisory processes and temporal organization of behavior. *Philosophical Transactions of the Royal Society, London* 351:1405–12. [HHHJK]
Shallice, T. & Warrington, E. K. (1970) Independent functioning of the verbal memory stores: A neuropsychological study. *Quarterly Journal of Experimental Psychology* 22:261–73. [RAM]
Sherman, J. C., Schweickert, J., Growdon, J. & Corkin, S. (1988) Sentence comprehension in Alzheimer's disease. Paper presented at the 18th Annual Meeting of the Society of Neuroscience, Toronto. [aDC]
Shiffrin, R. M. & Schneider, W. (1977) Controlled and automatic human information processing II: Perceptual learning, automatic attending, and a general theory. *Psychological Review* 84:127–90. [aDC]
Shulman, H. (1970) Similarity effects in short-term memory. *Psychological Bulletin* 75:399–414. [RLL]
Smith, S. (1989) *Syntactic comprehension in Alzheimer's disease.*Paper presented at the Academy of Aphasia conference, Santa Fe, New Mexico [aDC]
Snowdon, D. A., Kemper, S. J., Mortimer, J. A., Greiner, L. H., Wekstein, D. R. & Markesbery, W. R. (1996) Linguistic ability in early life and cognitive function and Alzheimer's disease in late life: Findings from the Nun Study. *Journal of the American Medical Association* 275:528–32. [aDC]
Squire, L. R. & Zola-Morgan, S. (1991) The medial temporal lobe memory system. *Science* 253:1380–86. [aDC]
Sternberg, S. (1969) The discovery of processing stages: Extension of Donder's method. *Acta Psychologia* 30:276–315. [KJF]
Stevens, K., Kempler, D., Andersen, E. & MacDonald, M. C. (1996) Preserved use of semantic and syntactic context in Alzheimer's disease. Poster presented at the Meeting of Cognitive Aging, Atlanta. [aDC]
Stine, E. A. L. (1995) Aging and the distribution of resources in working memory. In: *Age differences in word and language processing,* ed. P. Allen & T. Bashore. North-Holland. [SK]
Stowe, L. (1998) The cognitive function of the inferior frontal lobe in sentence comprehension. *Journal of Cognitive Neuroscience* S126. [rDC]
Stromswold, K., Caplan, D., Alpert, N. & Rauch, S. (1996) Localization of syntactic comprehension by positron emission tomography. *Brain and Language* 52:452–73. [aDC]
Sturt, P. & Crocker, M. W. (1996) Monotonic syntactic processing: A cross-linguistic study of attachment and reanalysis. *Language and Cognitive Processing* 11:449–94. [arDC]
Swinney, D. & Love, T. (1998) Language input rate as a parameter on structural processing. *Proceedings of the 11th Annual CUNY Conference on Human Sentence Processing, New Brunswick, NJ, March 1998.* [MW]
Swinney, D. & Zurif, E. (1996) Syntactic processing in aphasia. *Brain and Language* 50:225–39. [aDC]
Swinney, D., Zurif, E., Prather, P. & Love, T. (1996) Neurological distribution of processing resources underlying language comprehension. *Journal of Cognitive Neuroscience* 8:174–84. [aDC]
Tabor, W., Juliano, C. & Tanenhaus, M. K. (1997) Parsing in a dynamical system: A attractor-based account of the interaction of lexical and structural constraints in sentence processing. *Language and Cognitive Processes* 12:211–71. [arDC]
Tabossi, P., Spivey-Knowlton, M. J., McRae, K. & Tanenhaus, M. K. (1994) Semantic effects on syntactic ambiguity resolution: Evidence for a constraint-based resolution process. In: *Attention and performance XV,* ed. C. Umilta & M. Moscovitch. MIT Press. [NJP]
Tanenhaus, M. K., Spivey-Knowlton, M., Eberhard, K. & Sedivy, J. (1995) The interaction of visual and verbal information in spoken language comprehension. *Science* 268:1632–34. [PW]
Taylor, A. E., Saint-Cyr, J. A. & Lang, A. E. (1986) Frontal lobe dysfunction in Parkinson's disease. *Brain* 109:845–86. [aDC]
Tim, P. A. & Wingfield, A. (1993) Is speech special? Perception and recall of spoken language in complex environments. In: *Adult information processing: Limits on loss,* ed. J. Cerella, J. Rybash, W. Hoyer & M. L. Commons. Academic Press. [AW]
Tirre, W. & Pena, C. (1992) Investigation of functional working memory in the reading span test. *Journal of Educational Psychology* 84:462–72. [arDC]
Titchener, E. B. (1908) *Lectures on the elementary psychology of feeling and attention.* Macmillan. [AW]
Tomoeda, C. K., Bayles, K. A., Boone, D. R., Kaszniak, A. W. & Slauson, T. J. (1990) Speech rate and syntactic complexity effects on the auditory comprehension of Alzheimer patients. *Journal of Communication Disorders* 23:151–61. [aDC]
Towse, J. N. (1998) On random generation and the central executive of working memory. *British Journal of Psychology* 89(1):77–101. [JNT]
Towse, J. N. & Hitch, G. J. (1995) Is there a relationship between task demand and storage space in tests of working memory capacity? *Quarterly Journal of Experimental Psychology* 48A:108–24. [rDC, JNT]
Towse, J. N., Hitch, G. J. & Hutton, U. (1998) A re-evaluation of working memory capacity in children. *Journal of Memory and Language* 39:195–217. [JNT]
Trueswell, J. C., Tanenhaus, M. K. & Garnsey, S. M. (1994) Semantic influence on syntactic processing: Use of thematic information in syntactic disambiguation. *Journal of Memory and Language* 33:265–312. [aDC, NJP]
Turner, M. & Engle, R. W. (1989) Is working memory capacity task dependent? *Journal of Memory and Language* 28:127–54. [aDC, MJK]
Ullman, M. T., Corkin, S., Coppola, M., Hickok, G., Growdon, J. H., Koroshetz, W. J. & Pinker, S. (1997) A neural dissociation within language: Evidence that the mental dictionary is part of declarative memory, and that grammatical rules are processed by the procedural system. *Journal of Cognitive Neuroscience* 9:266–76. [ZB, rDC]
Vallar, G. & Shallice, T., eds. (1990) *Neuropsychological impairments of short-term memory.* Cambridge University Press. [aDC]
Waldstein, R. & Baum, S. (1992) The influence of syntactic and semantic context on word monitoring latencies in normal aging. *Journal of Speech Language Pathology and Audiology* 6:217–22. [aDC]
Walenski, M., Figueroa, S. & Swinney, D. (1997) The sentence memory task: A methodological investigation. *Proceedings of the 10th Annual CUNY Conference on Human Sentence Processing, Santa Monica, CA, March 1997.* [MW]
Walker, H. A. & John, E. E. (1984) Interference and facilitation in short-term memory for odors. *Perception and Psychophysics* 36:508–14. [RLL]
Wanner, E. & Maratsos, M. (1978) An ATN approach to comprehension. In: *Linguistic theory and psychological reality,* ed. M. M. Halle, G. Miller & J. Bresnan. MIT Press. [aDC]
Warrington, E. K. (1975) The selective impairment of semantic memory. *Quarterly Journal of Experimental Psychology* 27:637–57. [SAK]
Warrington, E. K., Logue, V. & Pratt, R. T. C. (1971) The anatomical localisation of selective impairment of auditory verbal short-term memory. *Neuropsychologia* 9:377–87. [RAM]
Waters, G. & Caplan, D. (1996a) The capacity theory of sentence comprehension: A reply to Just and Carpenter. *Psychological Review* 103:761–72. [aDC]
(1996b) Processing resource capacity and the comprehension of garden path sentences. *Memory and Cognition* 24:342–55. [aDC, AM]
(1996c) The measurement of verbal working memory capacity and its relation to reading comprehension. *Quarterly Journal of Experimental Psychology* 49:51–79. [arDC]
(1997) Working memory and on-line sentence comprehension in patients with Alzheimer's disease. *Journal of Psycholinguistic Research* 26:377–400. [aDC]
(submitted) The effect of a digit load on syntactic processing in subjects with high and low working memory capacity. [aDC]
Waters, G., Caplan, D. & Hildebrandt, N. (1987) Working memory and written sentence comprehension. In: *Attention and performance XII: The psychology of reading,* ed. M. Coltheart. Erlbaum. [arDC]
(1991) On the structure of verbal short-term memory and its functional role in sentence comprehension: Evidence from neuropsychology. *Cognitive Neuropsychology* 8(2):81–126. [arDC]
Waters, G. S., Caplan, D. & Rochon, E. (1995) Processing capacity and sentence comprehension in patients with Alzheimer's disease. *Cognitive Neuropsychology* 12:1–30. [aDC, AM]
Waters, G. S., Rochon, E. & Caplan, D. (in press) Task demands and sentence comprehension in patients with dementia of the Alzheimer's type. *Brain and Language.* [aDC]
Waugh, N. C. & Norman, D. A. (1965) Primary memory. *Psychological Review* 72:89–104. [RLL]
Wickelgren, W. (1965) Acoustic similarity and retroactive interference in short-term memory. *Journal of Verbal Learning and Verbal Behavior* 4:53–61. [RLL]
Williams, H. L., Beaver, W. S., Spence, M. T. & Rundell, O. H. (1969) Digital and kinesthetic memory with interpolated information processing. *Journal of Experimental Psychology* 80(3):530–36. [RLL]
Wright, R. E. (1981) Aging, divided attention, and processing capacity. *Journal of Gerontology* 36(5):605–14. [aDC]
Yngve, V. (1960) A model and a hypothesis for language structure. *Proceedings of the American Philosophical Society* 104:444–66. [SK]
Zatorre, R. J., Evans, A. C., Meyer, E. & Gjedde, A. (1992) Lateralization of phonetic and pitch discrimination in speech processing. *Science* 256:846–49. [rDC]
Zeki, S., Watson, J. D. G., Lueck, C. J., Friston, K. J., Kennard, C. & Frackowiak, R. S. J. (1991) A direct demonstration of functional specialisation in human visual cortex. *Neuroscience* 11:641–49. [KJF]
Zurif, E., Swinney, D., Prather, P., Solomon, J. & Bushell, C. (1993) An on-line analysis of syntactic processing in Broca's and Wernicke's aphasia. *Brain and Language* 45:448–64. [aDC]
Zurif, E., Swinney, D., Prather, P., Wingfield, A. & Brownell, H. (1995) The allocation of memory resources during sentence comprehension: Evidence from the elderly. *Journal of Psycholinguistic Research* 24:165–82. [aDC]