Pruning the Right Branch:

Working Memory and Understanding Sentences

by

Rose M. Roberts

B.A. Psychology, 1994

California State University, Los Angeles

SUBMITTED TO THE DEPARTMENT OF BRAIN AND COGNITIVE SCIENCES
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY IN COGNITIVE NEUROSCIENCE
AT THE
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

SEPTEMBER 1998

Signature of Author: ______________________________
Department of Brain and Cognitive Sciences
July 31, 1998

Certified by: ______________________________
Suzanne Corkin, Professor of Behavioral Neuroscience
Thesis Co-supervisor

____
Mary C. Potter, Professor of Psychology
Thesis Co-supervisor

Accepted by: ______________________________
Mriganka Sur, Chair
Department of Brain and Cognitive Sciences

Pruning the Right Branch: Working Memory and Understanding Sentences

by

Rose M. Roberts

Submitted to the Department of Brain and Cognitive Sciences
on July 31, 1998 in Partial Fulfillment of the Requirements for the
Degree of Doctor of Philosophy in Cognitive Neuroscience

ABSTRACT

An experiment was conducted to determine whether tests used to assess working memory in different disciplines (neuroimaging, psycholinguistics, neuropsychology) are highly correlated, and thus whether they are equivalent measures of a unitary underlying function. Scores on the different tests (N-back, reading span, backward digit span) did not correlate highly, and were predicted by measures of different hypothesized components of working memory. These results indicate that working memory is best conceived of as a system of multiple, interacting components that contribute to different aspects of task performance, rather than as a single, unified resource, and that currently popular tests of working memory cannot be used interchangeably to measure working memory.

A second experiment was conducted to examine the relation between sentence memory and working memory, and to determine whether memory for sentences is a function of the number of clauses in the sentence, or the number of new discourse referents. Subjects heard sentences of different lengths (2 – 5 clauses) and structures (relative clause, sentential complement, double object). Double object sentences contained one additional discourse referent per clause than the other two sentence types. If new discourse referents are the units of sentence memory, performance should be worse on double object sentences. If clauses are the unit of sentence memory, accuracy should be the same for all three sentence types. There were no reliable differences between double object sentences and the other two sentences types, indicating the clauses are the units of sentence memory. Subjects recalled 2-clause sentences highly accurately, and recalled 4-clause and 5-clause sentences poorly. There were large individual differences in the recall of 3-clause sentences. Over half of this variance was accounted for by individual differences in working memory. Measures of two hypothesized working memory components, the central executive and the short-term store, each accounted for independent variance in the sentence memory score.

Thesis Supervisors: Suzanne Corkin, Professor of Behavioral Neuroscience
Mary C. Potter, Professor of Psychology

TABLE OF CONTENTS

Acknowledgements

I would like to thank my advisors and committee members: Suzanne Corkin, Molly Potter, Ted Gibson, Nancy Kanwisher, and Earl Miller.

I was fortunate to have had the opportunity to work with some wonderful UROPs, who started out as advisees, and ended up as friends: Sarah Dash, Kevin Doyle, Joanna Gaunder, Chris Hooker, and Fenny Lin.

Many thanks to Brad Postle, for 3 a.m. pre-grant brainstorming sessions, Mark Snow, for turning our Theoretical ideas into reality, and Joe Locascio, for helping us to figure out what it all meant.

I want to thank my friends, who have provided support both moral and immoral: Andrea D'Avella, Loren Frank, Sherri Hitz, Carsten Hohnke, Cindy Kiddoo, Erica Klempner, and Gina Rendon;

My housemates at the Hub, especially Ioana Popescu, Cathy Leamy, Colin Clifford, and Quaid Morris, for not changing the locks this year;

My officemates, Dan Grodner and Tessa Warren, who never demanded a sterile work environment, and who laughed at me, not just with me;

Ann Doyle, who never knew what to expect on the other end of the phone line, but who continued to call me anyway;

Chris Moore, for finding the middle road, and for never letting me take myself too seriously;

Vicki Goodrow (and Zoe), for reminding me that there is more to life than graduate school;

And finally, I want to thank my parents, for continuing to claim me as their daughter even when I threatened to quit and come live in their guest room, and for their unflagging support and encouragement. This one's for you, Mom.

Chapter 1

Introduction:

Dissecting Working Memory

Memory is not unitary but depends on the operation of potentially independent, but typically interactive, components. One of the jobs of a cognitive neuropsychologist is to identify these component and indicate how they interact with each other.

- Morris Moscovitch (1992)

Introduction

This thesis presents a contribution to ongoing efforts to identify and explain the components of one mnemonic function, working memory. In order to understand a complex function like working memory, it is imperative to incorporate converging evidence from a variety of domains: experimental psychological studies of normal subjects, neuropsychological investigations of brain-lesioned subjects, functional neuroimaging studies of brain activity, and studies of neuronal activity and selective lesions in monkeys. Such an interdisciplinary approach has two fundamental requirements: First, that the function under study be clearly defined, and modeled in a way that makes testable predictions; and second, that there is a clear understanding of the relation between the function itself and the tasks used to assess the function.

Working memory is an important topic of study, because the ability to store and manipulate information is central to high-level cognitive functions, such as reasoning, planning, problem solving, and understanding language. The working memory construct has been invoked by researchers in a variety of domains: animal neurophysiology, developmental psychology, cognitive psychology, psycholinguistics, neuropsychology, and cognitive neuroscience. This centrality, however, also renders working memory a difficult subject to study, for two reasons: First, researchers studying non-primate animals (i.e., rats, pigeons; (Olton, Becker, & Handelmann, 1979), monkeys (Goldman-Rakic & Friedman, 1991), and humans (Baddeley, 1983) define the function differently. Second, researchers studying humans, who use Baddeley's (1983) definition of working memory as the ability to store and manipulate information needed for complex cognitive tasks, use different tests to assess it. These factors have made it difficult to compare results across studies, both within and across disciplines, and to draw conclusions about the cognitive components and neural substrate of working memory in humans.

The introduction of the concept of working memory has led to progress within the fields of cognitive psychology and psycholinguistics. An important goal of cognitive neuroscience, however, is to integrate information about the neural substrate of a behavior, obtained from neurophysiological, neuropsychological, and neuroimaging studies, into cognitive models of the behavior. The fact that researchers using different experimental techniques and subject populations often use different tasks to assess working memory capacity has proven to be a hurdle facing those who attempt to develop models of the cognitive and neural bases of working memory. Even within a discipline, where similar tasks are used, there are multiple versions of each task, and many parameters that vary across experiments and laboratories. It is rare to see discussions of how such differences in materials and procedures may affect results and conclusions. The experiment in Chapter 2 highlights the problem with using different tasks to assess working memory: The results showed that scores on tests used in different disciplines are not correlated with each other, and are predicted by scores on measures of different hypothesized components of working memory. This observation suggests that it is not possible to compare results across experiments, without taking into account the factors to which each test is sensitive.

Another area in which the introduction of the working memory construct has proved useful is in the study of the role of working memory in understanding language. It has been observed that lesions of left posterior parietal cortex result in a severely reduced short-term storage capacity (a digit or word span of 1 to 3 items). Span-impaired patients, however, have a relatively intact ability to understand language. This discovery challenged the commonly held assumption that the verbal short-term storage buffer is required for language processing (Caplan & Waters, 1990; Caplan & Waters, 1995; Howard & Butterworth, 1989; Martin, 1987; Martin & Romani, 1994; McCarthy & Warrington, 1987a; McCarthy & Warrington, 1987b; Saffran & Marin, 1975). The concept of working memory, with its emphasis on the simultaneous storage and manipulation of information, has offered a new approach to the study of the relation between language processing and memory. The experiments in Chapter 3 investigated the role of working memory in memory for sentences.

I. Animal Studies

A. Rats

In the literature on animal learning, "working memory" is used to refer to the ability to retain information across trials within a test session (Olton et al., 1979; Olton & Feustle, 1981). Working memory is distinguished from reference memory, the animal's between-session memory for the test apparatus and the task demands. The classic test of working memory uses the radial arm maze, an apparatus with a several arms radiating from a central starting point. In radial arm maze tasks, the ends of the arms are baited. The rat is placed at the center, and allowed to retrieve the food. Working memory is defined as the rat's ability to remember which arms it has visited (evidenced by avoidance of arms from which it has already retrieved the food). Working memory in rats is impaired following lesions of the hippocampal formation (Olton & Feustle, 1981; Olton, Wenk, Church, & Meck, 1988). This notion of working memory differs from the usage in human studies, where working memory refers to the on-line storage and manipulation of information (Baddeley, 1983). The working memory/reference memory distinction in studies of animal learning is analogous to the episodic/semantic memory distinction (memory for specific people and events vs. memory for general information) in studies of human long-term memory (Tulving, 1972).

B. Monkeys

In most monkey studies, working memory refers to the ability to hold information on-line across a brief delay (Goldman-Rakic & Friedman, 1991), analogous to short-term memory or short-term storage in human studies (Baddeley, 1983). Interest in the role of prefrontal cortex in working memory arose from observations that monkeys with lesions in the region of the principal sulcus (area 46) performed poorly on tests requiring the maintenance of information across brief delays (Passingham, 1993). For example, in the delayed response (DR) task, a monkey watches while one of two food wells is baited, and both are covered with opaque lids. Then, a screen is lowered, blocking the monkey's view of the food wells. Following a delay, the screen is raised and the monkey is allowed to retrieve the food from the baited location. Lesions to the principal sulcus region of dorsolateral prefrontal cortex (area 46) in monkeys produce severe impairments on this task.

Funahashi et al. (Funahashi, Bruce, & Goldman-Rakic, 1989) recorded from neurons in the principal sulcus while monkeys performed an oculomotor DR task with eight possible target locations. They observed neurons that exhibited delay-period activity. Many of these neurons were directionally-specific: They responded best when the movement was in one direction, and weakly when the movement was in other directions. These results suggest that prefrontal neurons maintain information about the target location across a delay (1-6 sec) in the absence of direct stimulation or movements. In complementary studies, Funahashi et al. (Funahashi, Bruce, & Goldman-Rakic, 1993) made selective lesions to the principal sulcus, and observed performance on the oculomotor DR task. They found that unilateral lesions led to a delay-dependent impairment in performance in the contralateral hemifield, with progressive worsening across the 8-sec delay. Wilson et al., (Wilson, O Scalaidhe, & Goldman-Rakic, 1993) presented evidence for a distinction between spatial and object working memory, with the former subserved by dorsolateral prefrontal cortex, and the latter by more ventral regions, mirroring the "what-where" division that has been observed throughout primate visual cortex (Mishkin, Ungerleider, & Macko, 1983).

Petrides has also addressed the question of whether specific regions of prefrontal cortex perform specialized functions. In selective lesion studies in monkeys, lesions to the dorsolateral prefrontal cortex resulted in impaired performance on non-spatial working memory tasks, a result that conflicts with the view that this area is specialized for spatial working memory (Goldman-Rakic & Friedman, 1991; Wilson et al., 1993). Lesions to the dorsolateral regions (areas 46 and 9) led to impairments on self- and externally-ordered monitoring tasks (monkeys had to select a single item from among three choices, then on the next two trials select different items without returning to previously selected items). These functions were spared following lesions to posterior dorsolateral cortex (areas 8 and 6). Lesions to this posterior region cause deficits on tasks of conditional associative learning (Petrides, 1991; Petrides, 1995). Further, lesions of ventrolateral prefrontal cortex (areas 45 and 47/12) impaired performance on spatial and nonspatial DR (Petrides, 1994). These results indicate that prefrontal cortex is not a homogenous region with respect to mnemonic processing, but instead that different regions contribute to different functions. Dissociations do not reflect a simple spatial vs. object dichotomy, however, but instead depend on the type of computations that must be performed on the contents of storage (Petrides, 1995).

Miller and colleagues have focused on the different executive functions performed by prefrontal cortex, rather than on functional differences among areas of prefrontal cortex (Miller, in press). They have studied three executive functions: selection of task-relevant information; integration of information from different processing streams (i.e., visual object and visual spatial information); and rule-learning. Recordings from neurons in the inferior convexity, ventral to the principal sulcus, revealed activity related to all three functions. For example, in one experiment monkeys were shown an array of objects, one of which was cued for later recall, and had to respond (after a 1.5 sec delay) if the object appeared in the same location as it had appeared before. Neurons showed task-specific responses: They were selectively active for specific objects, locations, or combinations of object and location. These neurons were hypothesized to mediate the selection of task-relevant objects, necessary for focal attention and response to the target. In another task where both object identity and spatial location had to be remembered, in turn, some neurons were selective for a specific object, when identity was relevant, and for a specific location, when location was relevant (Rao, Rainer, & Miller, 1997). Such neurons seem to have access to the information needed to integrate "what" and "where" information, and thus to form a coherent representation of an object's identity *and* its location in space.

Thus, while Petrides' studies of selective lesions in primates has focused on how different prefrontal areas perform different functions, Miller and colleague's work using neuronal recordings shows that neurons in the same area reflect (or can be trained to reflect) different task-relevant executive functions. Resolving the question of whether different areas perform different functions, or whether the same areas perform multiple functions depending on training and task demands, will require the convergence of data from selective lesion and recording studies.

Studies of working memory and executive function in monkeys have provided important clues to the functional organization of prefrontal cortex. However, the direct relevance of these studies to human working memory remains to be established. Several questions remain: First, while neurons in monkey prefrontal cortex appear to play a role in stimulus maintenance (i.e., DR) (Funahashi et al., 1989; Funahashi et al., 1993; Miller, Erickson, & Desimone, 1996; Passingham, 1993), humans with extensive lesions to the frontal cortex are not impaired on simple maintenance tasks (Milner & Petrides, 1984; Petrides & Milner, 1982). Second, monkeys are extensively trained, over months, on these tasks, using small sets of stimuli that are repeated and thus become familiar. It is not clear to what extent this extensive training may be tuning the neuronal activity, leading to the observed task-relevant responses. Human subjects, by contrast, can perform more complex tasks after only a brief verbal explanation, and can generalize to novel stimuli (i.e., spatial locations, objects, words). It has yet to be determined whether neuronal responses observed in monkeys after training (i.e. Miller, in press) have a direct analog in humans performing working memory tests.

II. Human Studies

A. Cognitive Psychology

Within cognitive psychology, the concept of working memory was introduced as an update to the idea of a passive short-term storage capacity, such as Miller's classic description of the magical number 7 +/- 2 as the capacity limit of immediate memory (Miller, 1956). The working memory theory was intended to include the idea of manipulations or computations being performed on the contents of storage, and thus to describe a more ecologically valid construct: the capacity-limited workspace used for sentence comprehension, reasoning, and problem solving (Baddeley & Hitch, 1974).

There are two main classes of working memory models; they differ in both the description of working memory, and in the tests used to assess its capacity. Multicomponent (MC) models hold that working memory consists of a set of interacting subsystems each dedicated to different components of the overall function. While there are some differences between specific MC theories, in general, these theories agree that there are subsystems dedicated to the storage of different types of information, and a central executive (CE) that is used to manipulate stored information and to coordinate the activity of the various subsystems (Baddeley, 1983; Baddeley & Hitch, 1974; Baddeley & Hitch, 1994; Smith & Jonides, 1997). By contrast, single resource (SR) models hold that working memory is best conceived as a pool of processing resources that can be flexibly deployed in the service of a wide variety of cognitive tasks. Processing resources are used for storage and for performing computations. For difficult tasks, a large amount of the available resources are required for processing, and thus fewer resources are available for storage. Individuals may differ in resource capacity, processing efficiency, or both (Daneman & Carpenter, 1980; Just & Carpenter, 1992; Just, Carpenter, & Keller, 1996a). One purpose of this thesis is to evaluate these models of the architecture of working memory.

Multicomponent (MC) Models

The original MC model was proposed by Baddeley and colleagues (Baddeley, 1996; Baddeley, 1983; Baddeley & Hitch, 1974). In Baddeley's model, working memory is made up of two subsystems dedicated to the storage of verbal and visuospatial information, and a CE that coordinates the activity of the subsystems, and controls attention allocation, goal monitoring, and inhibitory functions. The least-elaborated component of the model is the CE; evidence exists that it is made up of dissociable functions (Baddeley, 1996; Lehto, 1996). Baddeley (1996) has provided preliminary evidence for at least three separate functions of the CE: the ability to carry out two tasks simultaneously; to override automatic responses and switch retrieval plans; and to attend selectively. The verbal subsystem (the phonological loop) consists of two parts: the phonological input store, (accessed by subvocal speech or directly through auditory input), which holds information for about 2 sec; and the articulatory rehearsal process, which actively refreshes the contents of storage.

The original MC model has been fruitful in generating research in several different disciplines, including cognitive psychology, neuropsychology, and neuroimaging. It has been extended by other researchers, who have provided evidence for multiple visual stores and a multi-component central executive, and have discussed the interaction between working memory and long-term memory (Ericsson & Kintsch, 1995; Lehto, 1996; Logie, 1996; Smith & Jonides, 1997). A model that defines the components working memory, what functions they perform, and how they interact leads to an improved understanding of the relation between the brain and behavior: It makes the explicit, testable prediction that different brain lesions should lead to impairments on different components of working memory. Patients with such dissociations have been found and studied, the results are discussed in the next section.

Gathercole & Baddeley (1993) have hypothesized that the phonological loop and the CE make dissociable contributions to language comprehension. In their theory, the phonological loop is used to maintain a phonological record of sentences just heard or read, a record that acts as a backup to the normal on-line comprehension process (Baddeley & Hitch, 1994). Such a record would be particularly useful for understanding sentences that are initially understood incorrectly and must be reanalyzed (i.e., The cotton clothing is made of is grown in Mississippi), or that contain a long list of items to be remembered (i.e., Please go to the store and buy bread, milk, eggs, cheese, oranges, and spinach). The phonological loop is not assumed to be necessary for normal, first-pass sentence processing, due to the fact that patients with impaired short-term storage functions retain their ability understand a wide variety of sentence types, and show comprehension deficits only on long or syntactically complex sentences (Caplan & Waters, 1990).

According to Gathercole and Baddeley (1993), syntactic and semantic processing are functions of the CE; no real evidence has been presented to support this assertion, however. The fact that patients with frontal-lobe lesions who have functional impairments in executive functions do not typically have syntactic processing deficits argues against the idea that the CE is the site of syntactic processing, although such findings do not rule out the possibility that syntactic processing is carried out by one part of a multicomponent CE. Caplan & Waters (in press) offer a proposal that directly addresses the role of working memory in language comprehension. Rather than hypothesizing that first-pass language processing relies on the CE, they propose that there is an additional subsystem within working memory that is specialized for the interpretation of language, which they call the separate language interpretation resource (SLIR). They introduce an important distinction between two types of processing: interpretive processing, the assignment of syntactic and semantic structure to sentences; and post-interpretive processing, the use of interpreted sentences for other verbal tasks, such as answering questions, reasoning, and planning actions. This distinction formalizes, within the domain of language processing, the distinction suggested by Ericsson & Kintsch (1995):

> It is necessary, therefore, to differentiate the function of memory in generating cognitive states from its function in relating different states. In the former case, memory buffers contain intermediate results, which are significant for the formation of the cognitive state but irrelevant once it has been formed. In the other case we are talking about the storage and retrieval of cognitive end products. (p. 224)

With respect to language comprehension, interpretive processing generates the cognitive state (syntactically and semantically interpreted sentences), and post-interpretive processing stores and retrieves those sentences as demanded by the requirements of other tasks.

Single Resource (SR) Models

Researchers favoring SR models have a different conception of working memory and its relation to language comprehension. According to the SR model proposed by Just, Carpenter, and colleagues, working memory consists of a flexible pool of processing resources that can be used for performing computations, and for storing the intermediate and final products of those computations (Just & Carpenter, 1992; Just et al., 1996a; King & Just, 1991). With respect to sentence processing, the relevant computations include assigning syntactic structure to sentences and using that structure to decide what the sentence means. The relevant storage includes storing partially processed sentential elements, and storing parsed sentential elements for use in other verbal tasks (i.e., reasoning, planning, and drawing inferences). Thus, in SR models there is no distinction between interpretive and post-interpretive processing: Both functions are fulfilled by the same pool of resources. In addition, SR models hold that the same resources are used for storage and for performing computations. Thus, there is no distinction between storage functions that MC models ascribe to the phonological loop, and coordinating functions that MC models ascribed to the CE. "The theory deals with the resources used to support language comprehension computations, not the phonological buffer/articulatory loop of Baddeley's (1992) theory" (Just et al., 1996a, p.773).

This feature of SR models, the assumption that storage and processing depend on a common resource pool, is challenged by the neuropsychological evidence, discussed in the next section, that shows clear evidence of functional dissociations between storage and CE impairments, and between working memory and syntactic processing impairments. In one discussion, Just and colleagues dismiss this inconsistency with neuropsychological data by claiming that aphasic patients do have working memory deficits (Just et al., 1996a). But even if this claim is true, it does not account for the fact that there are patients with working memory impairments (in the CE, the phonological loop, or both) who are not impaired at interpreting language. If the same pool of resources is used in processing and storage, it should not be possible to observe such dissociations.

Just et al. (Just et al., 1996a; Just, Carpenter, Keller, Eddy, & Thulborn, 1996b) described an fMRI study that they claimed supports their proposal that the same pool of resources is used for sentence comprehension and for working memory. They compared brain activation in two conditions, one condition in which subject read sentences, and another in which they read sentences and maintained in memory the sentence-final word. They found that the read-only condition activated Wernicke's and Broca's areas, and that the read-and-maintain condition activated the same areas, along with additional voxels in Wernicke's area. Just et al. view these results as supporting their contention that "maintenance draws, in part, on the same resources as does sentence comprehension" (p. 774). They claim that their SR model makes two predictions: First, that the two conditions "should activate some of the same brain areas involved in sentence comprehension" and, second, that "the degree of activation in a brain area activated by both conditions should be greater in the read-and-maintain condition because the demand on the common resource pool should be greater" (Just et al., 1996a, p. 774). These predictions, however, are not unique to the SR model. MC models, and Caplan & Water's SLIR model, (contrary to Just et al.'s characterization) also predict overlap in brain areas activated in the two tasks, and greater activity in the read-and-maintain condition. Both tasks require subjects to read and make semantic verifications about sentences; thus, both should activate brain regions used for sentence interpretation and verification. The read-and maintain condition adds a requirement, demanding the subjects to perform the additional function of maintaining a list of words. This additional requirement to store words could explain the additional activation observed in Wernicke's area, an area that may overlap with the region damaged in span-impaired patients.

In order to show that sentence comprehension and working memory are using the same pool of resources, the tasks used to assess the two functions should not overlap: Sentence comprehension should be assessed by having subjects read and understand sentences, and working memory should be assessed by having subjects store and manipulate words. If such a study were performed, and the results showed entirely overlapping areas of activation in the in the two tasks, then one could claim that both tasks recruit the same pool of processing resources. That claim cannot be made based on the tasks used in Just et al. (Just et al., 1996a; Just et al., 1996b). Finding that the same brain region is active in more than one task is not surprising, and such an observation does not indicate that the brain area is performing the same computation in the two tasks.

B. Neuropsychology

Neuropsychological evidence for a multicomponent verbal working memory comes from patients with selective impairments in verbal short-term storage following lesions to the left inferior parietal cortex (Basso, Spinnler, Vallar, & Zanobio, 1982; Belleville, Peretz, & Arguin, 1992; Saffran & Marin, 1975), or patients with impairments in executive functions following lesions to prefrontal cortex (Petrides, 1996; Petrides & Milner, 1982). Patients with short-term storage impairments have a selective deficit in digit and word span (1 to 3 items), but normal long-term memory and relatively normal language comprehension abilities, at least for short, syntactically simple sentences (McCarthy, 1987a; McCarthy, 1987b; see Caplan & Waters, 1990 for a review). Patients with lesions of lateral prefrontal cortex have preserved digit and word span and normal language comprehension, but may have a range of executive and attentional deficits (Petrides, 1994; Petrides & Milner, 1982). Baddeley and Wilson (1988) described a patient with bilateral frontal-lobe lesions who showed a prototypical pattern of results: impaired attention and motivation, and difficulty in inhibiting responses; but normal digit span, and a normal recency effect in free recall, indicating sparing of phonological loop function. Baddeley and colleagues (Baddeley, Bressi, Sala, Logie, & Spinnler, 1991) have hypothesized that the primary working memory deficit found in patients with Alzheimer's disease (AD) is a deficit of executive functions: AD leads to a disproportionate impairment on dual-task experiments relative to impairments on single tasks. This pattern suggests that subjects with AD have difficulty coordinating simultaneous performance on two tasks, a putative function of the CE (Logie, 1996).

Like monkeys with selective lesions, humans with lesions to the lateral prefrontal cortex are impaired at tasks requiring monitoring within working memory (Petrides, 1996). For example, subjects have deficits in the self-ordered choosing task, which requires selection of one stimulus from among a set on each trial, until all stimuli in the set have been selected without repetition. Success on this task requires the maintenance and updating of a record of the responses made on each trial (Petrides & Milner, 1982).

Milner et al. (1985) reviewed evidence that the prefrontal cortex is necessary for the performance of tasks requiring memory of the temporal organization of events (i.e., the order of recent events and their frequency). Data from subjects with unilateral and bilateral excisions of parts of lateral prefrontal cortex suggest that the deficits reflect interference from information from preceding trials, rather than an inability to retain new information across a delay. Frontal cortex appears to be required for the time-marking process that permits the discrimination of recent and past events.

C. Neuroimaging

Much of the recent work in neuroimaging has been based on the MC model, and has attempted to separate and localize the different components of working memory (Awh et al., 1996; Braver et al., 1997; D'Esposito et al., 1995; Fiez et al., 1996; Paulesu, Frith, & Frackowiak, 1993; Petrides, Alivisatos, Evans,

& Meyer, 1993a; Schumacher et al., 1996; Smith et al., 1995). Results from neuroimaging studies are consistent with the neuropsychological data: PET and fMRI studies provide converging evidence that the network for verbal working memory includes the left-hemisphere posterior parietal cortex hypothesized to mediate verbal storage; Broca's area and supplementary motor areas hypothesized to mediate rehearsal; and bilateral dorsolateral prefrontal cortex hypothesized to mediate the manipulation of stored items and the coordination of concurrent tasks (D'Esposito et al., 1995; Smith et al., 1995; Smith & Jonides, 1997).

In a series of PET studies, Petrides and colleagues have shown that the self-ordered choosing task described above activates dorsolateral frontal cortex (area 46 and 9), while an associative learning task activates a more posterior region (area 8) (Petrides, Alivisatos, Meyer & Evans, 1993). Two order-monitoring tasks (generate numbers from 1 to 10, avoiding repetition; monitoring sequence of presented numbers from 1 to 10 for the missing number) led to activation of dorsolateral frontal cortex (Petrides, Alivisatos, Evans & Meyer, 1993). These results support the hypothesis derived from human and monkey lesions studies, that this region of dorsolateral frontal cortex is required for monitoring information within working memory.

In functional neuroimaging studies, a task that has been used extensively to measure brain activity associated with working memory is the N-back task. In this task, subjects are asked to monitor a string of stimuli, and to respond when a target is presented, with a target defined as an item that is the same as one that occurred n items ago, or "n-back". This task requires subjects to monitor items within working memory, to temporally tag incoming items and to rapidly update the contents of storage. Performance on this task activates dorsolateral prefrontal cortex (Awh et al., 1996; Braver et al., 1997; Schumacher et al., 1996; Smith & Jonides, 1997). This activation is most likely due to the task requirement of temporally tagging and monitoring the order of presented items (rather than simply recalling their content), a function shown by lesion studies to rely on frontal cortex (Milner et al., 1985).

III. Interdisciplinary approaches

The study of patients with brain lesions has been tremendously informative in the attempts to model the role of different brain structures in working memory. Neuropsychological studies, however, will necessarily leave some questions unanswered because of the size and heterogeneity of naturally occurring lesions, and because a lesion will never be confined to one hypothetical mnemonic component (Milner et al., 1985). For example, the question of whether the CE is made up of different functional components with distinct neural substrates will be difficult to address by studying patients with frontal-lobe lesions. This question, however, has been addressed using a combination of techniques.

Petrides and colleagues have studied the contributions of prefrontal cortex to memory, using lesion studies in monkeys, deficit-lesion correlations in humans, and PET studies of activity in the human brain (Petrides, 1995a; Petrides et al., 1993a; Petrides et al., 1993b; Petrides & Milner, 1982). These studies have provided converging evidence that the transient storage of information is carried out by various modality-specific and multi-modal regions of posterior cortex. Executive functions are subserved by prefrontal cortex, with different regions of lateral prefrontal cortex performing different functions: The dorsolateral prefrontal cortex monitors and manipulates information within working memory; and ventrolateral prefrontal cortex actively retrieves information from long-term memory (Petrides, 1994; Petrides, 1995a; Petrides, 1995b; Petrides, 1996). This model offers an explicit neurobiological representation of multicomponent models, including evidence for a multicomponent executive system, and an explanation of the interaction of working memory and long-term memory. An important feature of this approach is the use of the same, or very similar, tasks in monkey and human lesion studies and in human neuroimaging studies. Use of the same tasks allows the synthesis of results across subject populations and experimental techniques. This multidisciplinary approach is an encouraging example of cognitive neuroscience in action: building a model that answers a question of central interest in the study of human cognition, from data collected using different techniques and methods, in humans and nonhuman primates.

Chapter 2

Measuring Working Memory:
A Comparison of Some Common Tests

The mystery does not get clearer by repeating the question.

-Rumi, 1088

Introduction

Baddeley noted more than 20 years ago that a major impediment to understanding working memory and its relation to other cognitive functions is the fact that "there are no generally accepted working memory span measures, nor is it clear how one would validate candidate measures, other than by showing that they correlate with performance across a wide range of cognitive tasks" (Baddeley, Logie, Nimmo-Smith, & Brereton, 1985, p. 126). Nonetheless, researchers continue to use multiple tasks to measure working memory, tasks that differ markedly in structure, response requirements, and difficulty, and that may not have been validated by showing that they correlate with other putative measures of the function. Often, the preferred measure of working memory differs for different methods (behavioral studies, functional neuroimaging) and subject groups (young normal subjects, healthy older subjects, and subjects with neurological diseases), making cross-disciplinary comparisons difficult. Even within a field, where similar tasks are used (as in the case of the reading span measure used in psycholinguistic research), there are multiple versions of each task, and numerous variations in procedure and scoring that make cross-laboratory comparisons difficult.

This diversity among tests used to measure working memory would not be a problem if the tests were equivalent. Recent evidence, however, suggests that scores on commonly used working memory tests may not be highly correlated, and that the different tests may be differentially sensitive to processes such as storage, response time, rapid stimulus manipulation, and other executive functions (Dobbs & Rule, 1989; Lehto, 1996; Waters & Caplan, 1996b). In the present experiment, I assessed the correlations among several commonly used measures of working memory capacity, to test the hypothesis that the different tests are equivalent measures of working memory.

There are two different classes of working memory model: multicomponent (MC) and single resource (SR) models. In MC models, working memory is characterized as a set of separate, dissociable components that interact during the performance of complex cognitive tasks, with different components contributing to different aspects of task performance (Baddeley, 1983; Baddeley & Hitch, 1974; Baddeley & Hitch, 1994; Gathercole & Baddeley, 1993). In SR models, working memory is characterized as a unitary, limited capacity pool of processing resources, some concerned with storage and others with manipulation of information (Daneman & Carpenter, 1980; Daneman & Carpenter, 1983; Just & Carpenter, 1992; Just et al., 1996a; King & Just, 1991). In SR models, the same resources perform processing and storage functions, so that if a task has high demands for either, the resources available to the other function will be reduced.

In the standard MC model, proposed and extensively tested by Baddeley and colleagues (Baddeley, 1996; Baddeley, 1983; Baddeley & Hitch, 1994), working memory consists of two subsystems dedicated to the storage of verbal and visuospatial information, respectively, and a CE that coordinates the activity of the subsystems, and also controls attention allocation, goal monitoring, and inhibitory functions. The verbal subsystem (the phonological loop) consists of two parts: The phonological input store, (accessed by subvocal speech or directly through auditory input), which stores information for about 2 sec; and the articulatory rehearsal process, which actively refreshes the contents of storage. In MC models, differences among working memory tests may reflect the fact that tests measure different components. In this experiment, I examined the relation between different tests of working memory, and measures of two components relevant to performance on verbal working memory tests: short-term storage capacity (thought to reflect phonological loop function) and cognitive speed (thought to reflect, in part, the efficiency of the CE) (Richardson, 1996).

The original MC model has been fruitful in generating research in several different disciplines, including cognitive psychology, neuropsychology, and neuroimaging. It has been extended by other researchers, who have provided evidence for multiple visual stores and a multi-component CE, and who have discussed the interaction between working memory and long-term memory (Ericsson & Kintsch, 1995; Lehto, 1996; Logie, 1996; Smith & Jonides, 1997). A model that defines the components of working memory, what functions they perform, and how they interact leads to an improved understanding of the relation between the brain and behavior. It makes the explicit, testable prediction that different brain lesions should impair different components of working memory. Patients with such dissociations have been found and studied, and results have largely supported the model.

Neuropsychological evidence for a multicomponent verbal working memory comes from patients with selective impairments in verbal short-term storage following lesions to the left inferior parietal cortex (Basso et al., 1982; Belleville et al., 1992; Saffran & Marin, 1975) or patients with impairments in executive functions following lesions to the prefrontal cortex (Petrides, 1996; Petrides & Milner, 1982). Patients with short-term storage impairments have a selective deficit of digit and word span (1 to 3 items), but normal long-term memory and relatively normal language comprehension abilities, at least for short, syntactically simple sentences (McCarthy, 1987a; McCarthy, 1987b; see Caplan & Waters, 1990, for a review). Patients with frontal-lobe lesions have preserved digit or word span and normal language comprehension, but may have a range of executive and attentional deficits (Petrides, 1994). Baddeley and Wilson (1988) described a patient with a bilateral frontal-lobe lesion who showed a prototypical pattern of results: impaired attention and motivation, and difficulty in inhibiting responses; but normal digit span, and a normal recency effect in free recall, indicating sparing of phonological loop function. Baddeley and colleagues (Baddeley et al., 1991) have hypothesized that the primary working memory deficit found in patients with Alzheimer's disease (AD) is a deficit in executive functions: AD leads to a disproportionate impairment on dual-task experiments relative to impairments on single tasks. This pattern suggests that subjects with AD have difficulty coordinating simultaneous performance on two tasks, a putative function of the CE (Logie, 1996).

Much of the recent work in neuroimaging has been based on the MC model, and has attempted to separate and localize the different components of working memory (Awh et al., 1996; Braver et al., 1997; D'Esposito et al., 1995;

Fiez et al., 1996; Paulesu et al., 1993; Petrides et al., 1993a; Schumacher et al., 1996; Smith et al., 1995). Results from neuroimaging studies are consistent with the neuropsychological data: PET and fMRI studies provide converging evidence that the network for verbal working memory includes left posterior parietal cortex, hypothesized to mediate verbal storage; Broca's area and supplementary motor areas, hypothesized to mediate rehearsal; and bilateral dorsolateral prefrontal cortex, hypothesized to mediate the manipulation of stored items and the coordination of concurrent tasks (D'Esposito et al., 1995; Smith & Jonides, 1997).

In SR models, the same "resources" are used for storage and performing computations: There is no distinction between storage functions that MC models ascribe to the phonological loop and coordinating functions that MC models ascribed to the CE. This feature of SR models, the assumption that storage and processing depend on a common resource pool, is challenged by the neuropsychological evidence discussed above that shows clear evidence of functional dissociations between impairments in storage and executive functions. It is not clear how SR theorists would map the concept of a single pool of processing onto brain function, but it seems reasonable to expect that if the same pool of resources is used in processing and storage, it should not be possible to observe dissociations between storage and CE functions.

MC and SR models are associated with different sets of tasks used to measure working memory. In MC models, tasks may differ depending on what component of working memory is being assessed. Forward digit and word span are used to measure phonological loop function, and other more complex tasks requiring stimulus manipulation are used to measure executive functions (Baddeley, 1996; Case, Kurland, & Goldberg, 1982; Klapp et al., 1983; Lehto, 1996; Petrides et al., 1993a; Petrides, Alivisatos, Meyer, & Evans, 1993b; Petrides & Milner, 1982; Smith & Jonides, 1997). In functional neuroimaging studies, a task that has been used extensively to measure brain activity associated with working memory is the N-back task. In this task, subjects are asked to monitor a string of stimuli, and to respond when a target is presented, with a target defined as an item that is the same as one that occurred n items ago, or "n-back". This task requires subjects to monitor items within working memory, to temporally tag incoming items and to rapidly update the contents of storage (Smith & Jonides, 1997), quintessential executive functions that have been shown in studies of humans and nonhuman primates to rely on the dorsolateral prefrontal cortex (Awh et al., 1996; Braver et al., 1997; Petrides, 1991; Petrides, 1995a; Petrides, 1995b; Petrides, 1996; Petrides et al., 1993a; Petrides et al., 1993b; Petrides & Milner, 1982; Schumacher et al., 1996; Smith & Jonides, 1997). In the 2-back and 3-back tests most commonly used, the number of items that must be stored is well below the limited capacity of the phonological loop (generally around 7 items in college-age subjects). The task requires that subjects maintain an ordered representation of the last n items presented, compare incoming items to the appropriate stored item, and update the contents of storage on each trial. Thus, it seems plausible to hypothesize that N-back is a relatively pure measure of CE function, and that it may not be sensitive to the capacity of the phonological loop (i.e., scores on N-back may not correlate with simple digit and word span).

Other tasks have been hypothesized to tap the CE component of working memory directly, independently of the phonological loop. These tasks include the missing digit task used by Klapp and colleagues (Klapp, Marshburn, & Lester, 1983), in which 8 randomly ordered digits between 1 and 9 are presented; subjects have to report which digit between 1 and 9 was missing from the sequence. Performance on this test was not improved by rhythmic grouping of stimuli, and was not disrupted by irrelevant articulation, factors that affect phonological loop measures such as digit and word span. Lehto (1996) used a memory updating task, in which subjects were presented with lists of random digits of varying (but unpredictable) lengths, and asked to recall the last three or four digits of each list in order. Scores on this task were not correlated with scores on digit span or word span. Dobbs & Rule (1989) used a working memory task similar to N-back, in which randomly ordered digits were presented, and subjects were asked to report either the digit just heard, the digit one ago, or the digit two ago. Scores on this test did not correlate with storage capacity. These results indicate a dissociation between the storage and executive components of working memory, and show that these tests are measures of some component of executive function that is independent of short-term storage capacity.

In SR models, no distinction is made between the storage and executive functions of working memory: Working memory is the total pool of processing resources, some dedicated to performing computations, and others dedicated to storing the intermediate and final products of those computations (Case, Kurland, & Goldberg, 1982; Daneman & Carpenter, 1980; Daneman & Carpenter, 1983). The task most commonly used to assess working memory capacity is Daneman & Carpenter's (1980) reading span task, in which subjects read sets of sentences and then recall the final word of each sentence in the set. Set size is systematically increased, and reading span is defined as the largest set of sentences for which the subject can correctly recall all the final words on the majority of trials (generally 3 – 5 sentences). According to SR theorists, this task is interpreted as requiring the simultaneous storage and processing of information. Many different versions of the reading span task have been used, as well as other tasks with the same structure (subjects must simultaneously store and perform computations) but different processing requirements (i.e., sentence verification, performing math problems, counting dots, categorizing words) (Baddeley et al., 1985; Case et al., 1982; Engle, Cantor, & Carullo, 1992; Turner & Engle, 1989; Waters & Caplan, 1996a). There are significant correlations among different versions of the task, but even the most highly related measures share only about half of their variance, showing that there is a large effect of the specific processing requirement for each version of the test (Daneman & Merikle, 1996; Waters & Caplan, 1996b).

Correlations of $\underline{r}$ = .50 - .60 have been observed between different versions of reading span and word span, a test of short-term storage, indicating that reading span is sensitive to storage capacity (Daneman & Carpenter, 1980;

Light & A., 1985; Wingfield, Stine, Lahar, & Aberdeen, 1988). SR models, however, focus on individual differences in working memory capacity, including age related differences, which may be due to differences in storage capacity, processing efficiency, or both. Thus, in addition to storage, processing speed is seen as an important predictor of working memory capacity. SR models predict that subjects with slow or inefficient processing should have reduced working memory capacities (Engle et al., 1992). Reduced cognitive processing speed is the major determinant of age-related declines in scores on the reading span test (Babcock & Salthouse, 1990; Salthouse, 1990).

Waters & Caplan (1996) presented several criticisms of the traditional reading span task as a measure of working memory. First, they pointed out that the task does not require subjects to manipulate the contents of storage, a crucial aspect of many operational definitions of working memory. The storage and processing aspects of the task are relatively independent, and subjects may be making tradeoffs between the processing and recall task demands in different ways. Strategic differences may reduce the reliability of the test and make the results difficult to interpret. In addition, while the task requires processing and storage, only storage is measured. Waters & Caplan compared a number of different versions of the reading span task: They attempted to measure storage and processing by recording recall accuracy for the sentence-final words, and reaction time to sentence verifications. They found moderate-to-high correlations among the different versions of the task ($\underline{r}$ = .52 - .71), but found that correlations between the span measures and two other putative working memory measures (requiring subjects to generate random lists of numbers or shapes), were small or nonexistent (r = .07 - .32). Several factors were extracted from a factor analysis of these data, but no factor was extracted that represented a unitary limited-capacity working memory system that varied systematically across individuals.

While the reading span measure was initially intended to assess simultaneous storage and processing, the essence of working memory in SR theories, there are other possible interpretations of the task requirements. reading span, and its variants, are basically dual-task experiments, in which subjects are required to shift their attention between the processing and storage tasks. Switching attention, and coordinating performance on two tasks, are fundamental functions of the CE (Baddeley, 1996; Engle, 1996). Based on the observation of inconsistent effects of high and low memory span on a variety of tasks (see Engle, 1996 for a discussion), Baddeley (1996) has suggested that complex span measures, such as reading span, may be differentiating between subjects who are good and poor strategy users, rather than measuring some fundamental capacity limitation.

Thus, there are clear differences between different working memory tests: Some, such as N-back, are hypothesized to measure CE function independent of short-term storage capacity; while others, such as reading span and its variations, may measure other aspects of executive function but are also sensitive to storage capacity. If working memory (or at least CE component, Just & Carpenter, 1992) is a unitary, single resource, scores on different tests purporting to measure this resource should be correlated. A few studies have examined the relation among different working memory tests, and the results have generally not supported this prediction: Light & Anderson (1985) found a weak correlation between reading span and backward digit span (r = .34). Lehto (1996) found that backward digit span was not significantly correlated with memory updating or reading span (MU r = .32, RS r = .16), while memory updating and reading span were correlated (r = .57). Dobbs & Rule (1989) found that backward digit span was not correlated with a test similar to the N-back task (r = .14 - .27).

The results of the experiments reviewed above suggest that working memory is best conceived as being made up of multiple, interacting components with different properties, rather than as a single, unitary resource, and that different tests may measure these different components. In the present experiment, I sought to extend these findings by directly comparing subjects' performance on three commonly used working memory tests: N-back, backward digit span, and reading span. I also used a fourth working memory measure, category span, which is similar to reading span but incorporates a different processing task, requiring semantic categorization of words, rather than reading sentences. I chose this task for two reasons: First, to augment and validate the reading span measure, by including a task with the same requirements of shifting between two tasks, but with a different background processing requirement. Using multiple measures is important to ensure that results are not due to idiosyncrasies of the materials on any given task (Engle et al., 1992; Turner & Engle, 1989). Second, the processing requirement of category span (to report the category of a set of words), requires subjects to search and retrieve from semantic memory, and thus to require on an additional putative executive function. Baddeley and Wilson (1988) described a patient with a bilateral frontal lobe lesion and impairments on a wide range of CE functions. In addition to his other deficits, he was impaired at generating words from a given semantic category. In the experiment reported in Chapter 3 I found a high correlation (r = .65) between reading span and category span, and also found that category span correlated more highly with a measure of sentence memory capacity than did Reading Span.

I next examined the relation between the four working memory tests and two factors that may contribute to working memory capacity, simple storage ability and cognitive processing speed. In MC models, these two factors reflect the capacity of different working memory components, the phonological loop and the CE. Both models predict that working memory measures should be correlated with measures of processing speed: MC models because speed reflects the efficiency of the central executive (Richardson, 1996); and SR models because fast and efficient information processing leaves more resources available for storage (Just & Carpenter, 1992). Reading span, category span, and backward digit span, tests that require subjects to store as many items as possible, should be correlated with storage capacity. N-back, however, is a measure of the ability to rapidly manipulate and update stored information; it does not require subjects to store large numbers of items, and thus may not correlate with storage capacity. If these predictions are correct, then any attempts to compare results across experiments without taking into account the factors to which each test is sensitive will be misleading and uninformative. If scores on the different working memory tests (a) do not correlate with each other and (b) are predicted by measures of different component variables, this finding would argue against the idea the working memory is best understood as a unitary, single resource, contradicting the claims of SR models.

Methods

Subjects

We tested 60 MIT undergraduate and graduate students (age range 18-32 years, mean age 20.2) in two 1-hour experimental sessions on two different days. Subjects were paid for their participation.

Materials

Working Memory Tests

N-back. Subjects saw words (4-letter abstract nouns) presented one at a time on the computer screen at the rate of one word every 3 sec (2500 ms word presentation, 500 ms interstimulus interval). They responded with a button press whenever they saw a target. A target was defined as a word that was the same as the word presented N ago, or "N-back." Subjects were first presented with 2-back targets; if they reached criterion (70% correct) they were presented with 3-back targets, then 4-back, and then 5-back. There were 70 - 80 trials at each set size, with 10 correct targets (hits) per set. The score for each set size was computed by subtracting the number of false alarms from the number of hits to correct for guessing; then scores for all levels completed were combined to reach a composite N-back score. The equation for combination was as follows: 1 + ((2-back, % correct) + (3-back, % correct) + (4-back, % correct) + (5-back, % correct) X 100).

Backward Digit Span (bDS). Subjects heard strings of digits presented at the rate of 1 digit per sec, and then recalled them in reverse order. Span was defined as the longest string of digits a subject correctly repeated in reverse order on at least one of two trials.

Reading Span. Subjects viewed sets of short declarative sentences (5-10 words, mean 7.3) on the computer screen, and read them aloud. Next, subjects viewed simple questions (probing either the subject or the main verb) and answered them aloud. After two sentence-question sets, subjects were prompted to recall the final word of both sentences. Subjects were first presented with five trials at set-size two; in order to advance to larger set sizes (three to six), they had to recall all the words correctly on three of the five trials. Span was defined as the largest set size at which subjects recalled all of the words correctly on four of the five trials; with an additional .2 added for each trial they recalled correctly at the next set size.

Category Span. The procedure was the same as for Reading Span, except that subjects read a list of four nouns, three of which belonged to a common category (i.e., animals, foods, or colors). The fourth word did not match the category. On each trial, subjects reported the category name aloud. After two such trials, subjects recalled the mismatch word for the two lists. If the subject recalled the two words correctly on three of five trials at set size two, the set size was increased to three. The largest possible set size was six. Scoring was the same as for Reading Span.

Short-term Storage Tests

Forward Digit Span (fDS). Subjects heard strings of digits presented at the rate of 1 digit per sec, and then recalled them in order. Span was defined as the longest string of digits a subject could repeat correctly, in order, on one of two trials.

Word Span (WS). Subjects heard lists of words (1-syllable concrete nouns) presented at the rate of 1 word per sec, and then recalled them in order. Span was defined as the longest string of words a subject could repeat correctly in order on one of two trials.

Cognitive Speed Tests

Choice Reaction Time (Choice RT). Two words, "push" and "rest," were presented, one on the left side and one on the right side of the computer screen. Subjects looked for the word "push" and pressed the left button on a box if "push" appeared on the left, and the right button if "push" appeared on the right. "Push" appeared on the left on half of the 100 trials. The ISI varied randomly between 500 and 2500 ms. Subjects were instructed to respond as quickly as possible while still being accurate.

Go/No-Go Decision Time (GNG-DT). The words "move" or "stay" were presented on the screen while subjects held down one button of a two-button box. If the word "move" appeared, subjects released the first button and hit the second button as quickly as possible. Decision time was the time that elapsed between word presentation and the release of the first button. There were 100 trials; "move" was presented on 80% and "stay" on 20% of the trials. Three ISIs were used: 500, 750, and 1000 ms.

Go/No-Go Movement Time (GNG-MT). The task was the same as above. Movement time was the time that elapsed between releasing the first button and pressing the second button.

Digit-Digit Matching (DigDig). Pairs of digits were presented at the center of the screen. Subjects pressed a key marked "S" if the two digits were the same, and a key marked "D" if they were different. The digits were the same on half of the 90 trials. Subjects were instructed to respond as quickly as possible while still being accurate.

Digit-Symbol Matching (DigSym). A key consisting of the digits 1 to 9, each paired with an abstract symbol, was presented across the top of the screen. At the center of the screen, a digit-symbol pair was presented, and subjects had to decide whether the test pair was the same as the pair in the key. Subjects pressed a key marked "S" if the digit-symbol pair was the same as the pair in the key, and a key marked "D" if the pairing was different. Each digit-symbol pair was probed 10 times, for a total of 90 trials. The digit-symbol pair was the same on half of the trials. Subjects were instructed to respond as quickly as possible while still being accurate.

Procedure

On the first day of testing, subject were given the following tests: Forward Digit Span, Backward Digit Span, Word Span, Reading Span, Category Span, and N-back. Because superior performance on the working memory tests resulted in longer testing sessions, half the subjects received the tests in the above order, while for the other half N-back was tested before Reading Span. This manipulation allowed us to determine whether fatigue affected performance on the working memory tests. On the second day of testing, subjects were given the remainder of the tests in the following order: Choice RT, Go/No-Go, Digit-Digit, and Digit-Symbol.

Data Analysis

We used three types of statistical analyses: correlation, multiple regression, and factor analysis. For the correlations, a Bonferroni correction was performed on the p values in order to control for the increased Type 1 error associated with multiple significance tests. Given the number of comparisons conducted, a p value of $p < .001$ was required for a correlation to be considered significant. We calculated stepwise multiple regression analyses of each working memory variable on the set of storage and cognitive speed variables. The entry and removal criterion was $p = .1$. We also performed an oblique promax factor analysis on the pooled set of all working memory and predictor variables. Oblique factor analysis differs from orthogonal factor analysis by allowing the factors to be correlated. The number of factors extracted was based on Kaiser's stopping rule, which specifies that only those eigenvectors with eigenvalues of at least 1 should be retained in the model (Bryant & Yarnold, 1995). Squared multiple correlations of each variable with every other variable were used as the prior communality estimate (Harman, 1976).

Results

Means, standard deviations, and ranges are presented for each of the working memory tests, and for the seven component variables, in Table 1. A t-test of the order variable (Reading Span first vs. N-back first) showed no significant order effect. Four scores, for three subjects, two on Choice RT, one on GNG-DT, and one on Digit-Digit Matching, were more than three standard deviations from the group mean; these four scores were replaced with the mean plus three standard deviations. None of the variables had distributions that deviated markedly from normality except for Choice RT, GNG-DT, and GNG-MT, which showed positive skewing. Because the significance tests reported below assume normality, a log transformation was performed on those three variables, after which they achieved normality.

Table 1. Descriptive Statistics for Working Memory and Component Tests

Test	Mean	SD	Range
N-back	3.2	0.8	1.5 - 4.7
Backward Digit Span	6.7	1.8	3.0 – 10.0
Reading Span	3.9	1.1	2.2 - 6.0
Category Span	2.6	1.0	1.2 - 4.6
Forward Digit Span	8.0	1.3	5.0 – 10.0
Word Span	6.3	1.1	4.0 - 9.0
Choice Reaction Time	423	65	315 – 648
Digit-Digit Matching	565	66	439 – 786
Digit-Symbol	1020	151	731 – 1449
Go/No-Go Decision	323	53	244 – 492
Go/No-Go Movement	109	29	59 – 191

Correlations between the working memory tests are presented in Table 2. Significant correlations were observed between N-back and Backward Digit Span, and between Reading Span and Category Span. No other correlations were significant at the p < .001 level; however the correlation between Backward Digit Span and Reading Span (r = .36) was significant at p < .005. N-back was not significantly correlated with Reading Span or Category Span (ps < .11).

Table 2. Correlations Among Scores on Working Memory Tests

	BDS	Reading Span	Category Span
N-back	**.41***	.21	.20
Backward Digit Span	-	.36	.24
Reading Span		-	**.57***

* p < .001

Correlations between the working memory and the storage and speed variables are presented in Table 3. The storage variables, measured by Forward Digit Span and Word Span, correlated significantly with Backward Digit Span. Forward Digit Span and Word Span were significantly correlated (r = .63). Word Span correlated with Reading Span (r = .53, p < .0001) and to a lesser extent with Category Span (r = .37; p < .004). N-back, however, was not significantly correlated with either storage variable at p < .001; the correlation between N-back and Forward Digit Span (r = .33) was significant at p < .01, and the correlation with Word Span was not significant (p < .14). The correlations between the working memory variables and the speed variables showed the opposite pattern of results. The only significant correlations at p < .001 were between N-back and two of the speed measures, Choice RT and Digit-Digit

Matching. N-back was correlated with the rest of the speed measures to a lesser degree: Digit Symbol Matching (r =.29, p < .02); GNG-Decision Time (r = .37, p < .004) and GNG-Movement Time (r = .40, p < .002). Backward Digit Span was moderately correlated with two of the speed measures, Choice RT (r = .30, p < .02) and GNG-Decision Time (r = .34, p = .008). The correlations between Reading Span, Category Span, and the speed measures did not approach significance.

Table 3. Correlations Between Working Memory and Short-Term Storage and Speed Tests

	N-back	BDS	Reading Span	Category Span
Forward Digit Span	.33	**.59***	.26	.29
Word Span	.19	**.55***	**.53***	.37
Choice RT	**.46***	.30	.06	.09
DigDig	**.43***	.25	.15	.02
DigSym	.29	.08	.08	.05
GNG-DT	.37	.34	.09	.16
GNG-MT	.40	.25	.12	.05

* p < .001

The results from the stepwise multiple regression analyses are presented in Table 4, which shows, for each working memory test, the percentage of variance accounted for by the set of retained predictors, and the p values for each of these predictors. Only the four predictor variables listed in the table were significantly related to the working memory variables: Forward Digit Span, Word Span, Choice RT, and Digit-Digit Matching. Of the variance on Backward Digit Span, 40% was accounted for by the linear combination of Forward Digit Span and Word Span. Of the variance on N-back, 27% was accounted for by the combination of Choice RT and Digit-Digit Matching. Word Span was the only significant predictor of Reading Span ($R^2 = .28$) and Category Span ($R^2 = .13$).

Table 4. Multiple Regression Analyses for Working Memory Tests: Percentage Variance Accounted for and p Values for Significant Predictors

	R^2	fDS	Word Span	Choice RT	DigDig
Backward Digit Span	.40	.004	.029	n.s.	n.s.
N-back	.27	n.s.	n.s.	.015	.056
Reading Span	.28	n.s.	.0001	n.s.	n.s.
Category Span	.13	n.s.	.004	n.s.	n.s.

In order to explore the underlying relations among these variables, an exploratory factor analysis was performed. Two factors were extracted based on Kaiser's stopping rule, (retain only those eigenvectors with eigenvalues of at least 1) (Bryant & Yarnold, 1995). After the promax rotation, the correlation of the two factors was .40. The variance of the two factors (variables standardized) were 3.55 and 2.84 (2.61 and 1.9 when each factor was adjusted for the other). A variable was considered to load on a factor if it had a factor loading of .3 or higher (Bryant & Yarnold, 1995). Each variable loaded on only one of the two factors. All the speed variables as well as the N-back loaded on Factor 1, while the storage variables and the other three working memory measures loaded on Factor 2. Factor loading coefficients of each variable on both factors are shown in Table 4.

Table 4. Factor Loading Coefficients for Each Variable

Test	Factor 1	Factor 2
Forward Digit Span	.23	.61
Word Span	.09	.74
Backward Digit Span	.14	.61
Reading Span	-.15	.72
Category Span	-.13	.62
N-back	.47	.19
Choice Reaction Time	.75	.01
Digit-Digit Matching	.80	-.05
Digit-Symbol Matching	.75	-.16
Go/No-Go Decision Time	.73	.12
Go/No-Go Movement Time	.68	0

Figure 1 shows a path diagram of the factor loadings for each variable. In the path diagram, the straight arrows indicate a causal effect of the factor on the variable to which it points. The path coefficient associated with each arrow indicates how much the variable changed (in units of its SD) when the factor changed one unit of its own SD. The curved arrow indicates a correlation between the two factors whose causal basis is not explicated in the model.

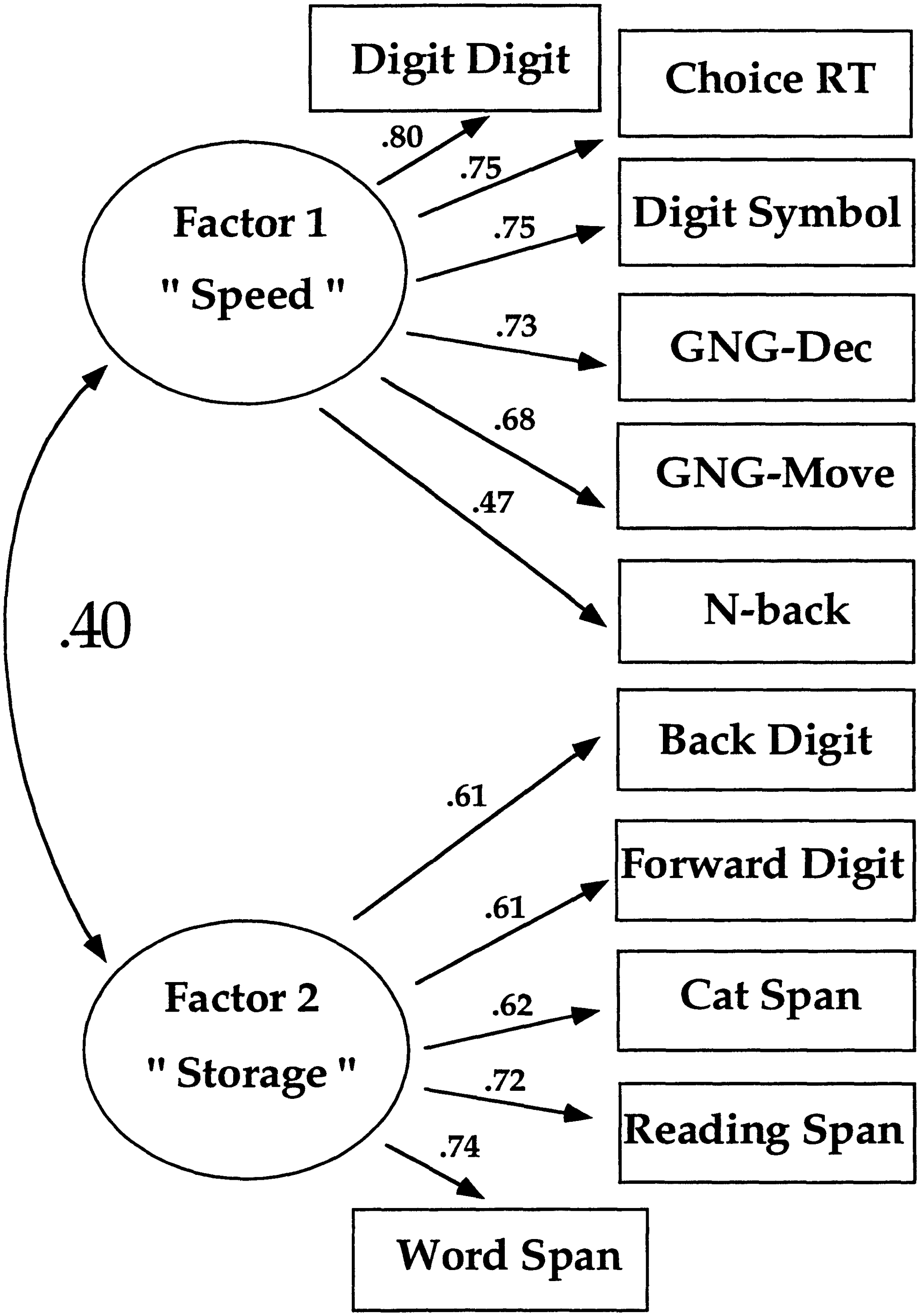

Figure 1. Path diagram of the factor loadings for each variable.

Discussion

I examined the relation among different commonly used working memory tests, and found moderate to nonsignificant correlations among them. Using the strict criterion suggested by a Bonferroni correction ($p < .001$), only N-back and Backward Digit Span, and Reading Span and Category Span, were significantly correlated. Using a more lenient criterion ($p < .05$), Backward Digit Span and Reading Span were also significantly correlated. Under no criteria were N-back and Reading Span or Category Span significantly correlated. Further, the working memory tests were correlated with different predictor variables: Backward Digit Span and Reading Span were correlated with storage capacity, whereas N-back was correlated with measures of cognitive speed. With a lenient criterion, N-back was related to one of the two measures of storage capacity, Forward Digit Span (but not Word Span), while Backward Digit Span was correlated with two speed measures. Again, under no criteria were Reading Span or Category Span significantly correlated with any speed variables. The results of the multiple regression and factor analysis helped clarify the relations among these variables: Multiple regression showed that for N-back, unique variance was explained only by two of the speed variables; and for Backward Digit Span, Reading Span, and Category Span, unique variance was explained only by the storage variables, suggesting that the correlations that might be considered significant under a more lenient criterion than $p < .001$ are, in fact, spurious. Factor analysis showed two moderately correlated factors underlying the set of variables tested, one of which could be interpreted as a "speed" factor and the other a “storage” factor. N-back loaded on the first factor, while the other three tests loaded on the second factor, providing further evidence that these tests measure dissociable aspects of working memory.

The N-back task is not sensitive to a subject’s storage capacity, and is thus not a measure of the short-term storage component of working memory. Because of the task demands (subjects must integrate and temporally tag incoming words, and compare them to the appropriate stored word, then update the contents of storage) and because of the correlation with the cognitive speed tests, it is reasonable to hypothesize that N-back is a measure of the rapid updating and manipulation of information within working memory, and thus a relatively pure measure of at least some aspects of executive function. It is unlikely that N-back loaded on the speed factor simply because test items were presented at a fixed rate, rather than being contingent upon subject’s responses, because the presentation rate was slow (1 word every 3 sec). Subjects were not instructed to respond as quickly as possible, but were instructed simply to respond within the 3 sec period. Subjects (who had simple reaction times on the order of 300 ms) had no difficulty in making responses within this period, and did not report any difficulty with this task due to the rate of stimulus presentation.

The hypothesis that N-back measures CE functions is consistent with neuroimaging results showing significant activation in dorsolateral prefrontal cortex in 2- and 3-back experiments, when the temporal ordering and stimulus manipulation demands are greatest, but not in 0- and 1-back conditions (Smith & Jonides, 1997). FMRI studies have also demonstrated that increasing the difficulty of N-back by increasing N from zero to three leads to increased activation in a fixed set of brain regions, rather than the recruitment of new areas, a finding that is consistent with the idea that working memory consists of a fixed number of basic components (Smith & Jonides, 1997).

What are the task requirements of N-back that give rise to the activation observed in dorsal prefrontal cortex? N-back requires more than memory for content: Subjects must remember the order in which items are presented, and must update the contents of storage on each trial. In humans and monkeys, dorsolateral prefrontal cortex is critical when an experimental task requires manipulation of stored information, even when the number of items to be manipulated is less than the subject's storage span (Petrides, 1995, 1996; Petrides & Milner, 1982; Petrides, Alvisatos, Evans & Meyer, 1993; Petrides, Alivisatos, Meyer & Evans, 1993). N-back requires more than simply maintaining a record of the last "n" items presented: The crucial requirement for success on the N-back task is remembering the order of the stored items, so that new stimuli can be matched to the appropriate stored item. Memory for temporal order of events is a function subserved by prefrontal cortex: Milner et al. (1985) have shown that subjects who underwent unilateral frontal lobectomy were impaired at judging which of two items they saw more recently, and in judging how frequently they have seen stimuli. These subjects with temporal ordering deficits were unimpaired at simple recognition tests with the same stimuli. Subjects who underwent unilateral temporal lobectomy, by contrast, were impaired at content recognition but unimpaired at making recency and frequency judgments. Thus, memory for content and memory for temporal order are doubly dissociated in brain-lesioned subjects, evidence that the functions are subserved by separate neural systems.

Performance on the Reading Span test, in contrast to N-back, and in conflict with predictions of SR models, is not influenced by cognitive speed in young subjects. Instead, in agreement with previous research (Daneman & Carpenter, 1980,1983; Light & Anderson, 1985; Wingfield et al., 1988) Reading Span was correlated with Word Span, a measure of simple storage ability. The lack of correlation between Reading Span and N-back, and the low correlation between Reading Span and Backward Digit Span suggests that if, in fact, Reading Span is measuring some aspect of executive function, it is a different aspect than that measured by N-back. One possibility is that one or more of the above tests is simply a poor or unreliable measure of working memory, as has been argued for Reading Span (Baddeley, 1996; Waters & Caplan, 1996b); another possibility is that different tests measure distinct executive functions. For example, it may be that N-back primarily measures updating within working memory, or memory for temporal order (Smith & Jonides, 1997), while Reading Span measures attention shifting (Engle, 1996) or efficient strategy use (Baddeley, 1996). Whichever interpretation is correct, the lack of correlations among different tests provides clear evidence against SR models that characterize working memory as a single, unified pool of processing resources. MC models have proven more useful for the purpose of testing brain-behavior relations: They allow researchers to make testable predictions about the number, nature, and interaction between working memory components, and the integration of neurophysiological data with cognitive models (Smith & Jonides, 1997).

There are some differences between the results reported here and previous attempts to compare working memory tests. Lehto (1996) used a Memory Updating task (described in the introduction) that he claimed independently measured CE function. He found that scores on Memory Updating correlated with scores on Reading Span, but not Backward Digit Span. This result is somewhat surprising given the present finding that N-back, which shares some similar task demands with Memory Updating, correlated with Backward Digit Span, but not Reading Span. However, this disagreement may be due to the fact that these tests all measure some overlapping but not identical aspects of CE function. The current finding, that N-back was the only working memory test that was sensitive to cognitive speed, suggests that rapid stimulus manipulation is an important determinant of performance on this test. However, speed tests such as the ones used here can be considered only indirect indices of the efficiency of CE functions (Richardson, 1996). An interesting follow-up experiment that would answer some of the questions raised by this study regarding what aspects of central executive function different working memory tests are measuring would be to look at the relation between these tests and some simple tasks designed to tap different hypothetical CE functions directly, such as selective or sustained attention, inhibition, or set-shifting. Clearly, an important goal for future research will be to define and explain the properties of the CE, the component of working memory that Baddeley (1983) called "the area of residual ignorance" (p. 315).

Conclusions

These results highlight the fact that tasks used to assess working memory should be pretested and validated, and that researchers must show the connection between the task they use as a measure of working memory, and the function itself. This connection may be demonstrated by presenting an analysis of the task demands, or by showing that the test is sensitive to the components of working memory that it is intended to measure. It would be extremely useful if researchers would perform such validating analyses, and then adopt some standard versions of tasks that could be used in different laboratories and with different subject groups. Without such validation and replication, it will be difficult to develop a model of working memory that will be broadly accepted outside of the laboratory in which it was generated.

The fact that different working memory tests are not highly correlated, and that they are correlated with measures of different component variables, suggests that working memory is best conceived of as a set of interacting components that are used to different degrees depending on the demands of the particular cognitive task. Rather than using complex tasks with multiple demands, a more useful approach to the study of working memory would be to design simple tests of the different hypothesized component processes. These tests could then be used to assess patterns of brain activation during functional neuroimaging, and patterns of impaired and spared test performance in patients with neurological disease. These data would then be appropriate for comparison across experimental methods and subject groups, and would aid researchers in building and refining models of the cognitive and neural bases of working memory.

Chapter 3

Working Memory and Remembering Sentences

These shifting and confused gusts of memory
never lasted for more than a few seconds.

-Proust

Introduction

This chapter addresses the question of the relevant processing unit in understanding and remembering sentences, and examines the relation between working memory and memory for sentences. Caplan and Waters (1998) distinguish two types of linguistic processing: interpretive processing, the assignment of syntactic and semantic structure to sentences; and post-interpretive processing, the use of interpreted sentences for other verbal tasks, such as answering questions, reasoning, and planning actions. Historically, the clause has been considered the central processing unit for on-line sentence comprehension (interpretive processing) and for off-line sentence memory (post-interpretive processing) (Tanenhaus & Trueswell, 1995). However, recent evidence examining interpretive processing of sentences has challenged this idea, and has suggested that new discourse referents (agents or events), rather than clauses, are the relevant units for interpretive processing (Gibson, 1998). This revised theory of interpretive processing suggests that it would be appropriate to reanalyze the data underlying the belief that the clause is the unit of post-interpretive processing. A review of the literature shows that the data used to support this assumption are open to alternative interpretations: Specifically, there are a number of confounds in previous studies, and consequently the data do not unambiguously support the idea that the clause is the unit of sentence memory. This experiment was designed to determine whether the clause is the unit of sentence memory, or whether, as in interpretive processing, sentence memory might be a function of the number of new discourse referents or content words in the sentence.

In measures of on-line and off-line sentence processing, researchers have observed large individual differences in performance (speed and/or accuracy of responses). The most common explanation for this variance is in terms of individual differences in working memory, the ability to store and manipulate information used in complex cognitive tasks (Baddeley, 1986; Daneman & Carpenter, 1980, Just & Carpenter, 1992). An alternative view, recently proposed by MacDonald & Christiansen (1998), is that working memory used for language processing does not exist: Individual differences are due to differences in reading skill and experience with language. The second goal of this study was to determine whether individual differences in working memory, independent of reading skill, can explain variance in sentence memory.

The Unit of Sentence Processing

Unit of interpretive processing

An early proposal regarding the unit of interpretive processing was made by Kimball (1973). He proposed that sentence comprehension was clause-based, such that at most two partially processed clauses could be maintained in working memory at one time (cf. the more recent related proposals of Stabler, 1994, and Lewis, 1996). This proposal accounted for the difficulty associated with processing *nested* structures such as (1):

(1) $[_S$ The student [who $[_S$ the professor [who $[_S$ the scientist collaborated with]] had advised]] copied the article].

A syntactic category A is said to be nested (or *center-embedded*) within another category B if B contains A, a constituent to the left of A, and a constituent to the right of A. In (1), the relative clause (RC) "who the professor . . . had advised" is nested within the sentence "the student . . . copied the article". Furthermore, a second RC "who the scientist collaborated with" is nested within the first embedded sentence "the professor . . . had advised". By contrast, left- or right-branching structures are much easier to understand than nested structures. For example, the right-branching structure in (2) has the same meaning as its nested counterpart in (1) at the level of thematic structure, but it is much easier to understand:

(2) The scientist collaborated with the professor who had advised the student who copied the article.

Kimball's clause-based proposal accounts for the contrast between nested and right-branching RCs as follows. Processing the first subject "the student" in (1) causes the initiation of a new clause that will be completed when the verb "copied" and its immediate dependents are located in the input string. Processing the following two subjects of the embedded RCs ("the professor" and "the scientist") causes the initiation of two further clauses, resulting in a total of three partially processed clauses, which is more than the proposed resource capacity. By contrast, there is never more than one incomplete clause while processing the right-branching structure in (2), so this sentence is processed without difficulty.

Although the clause-based proposal accounted for some nesting complexity effects, recent research has suggested that the difficulty that people have in processing nested structures is not because of a clause-based processing mechanism. Rather, nesting complexity seems to depend on two factors: 1) the number of syntactic heads that are required to form a grammatical sentence from a partially processed input string; and 2) the number of new discourse referents that have been processed since each required head was first known to be required (Gibson, in press). According to this theory, the point of maximal memory complexity in processing the sentence in (1) occurs at the point of processing the most embedded subject "the scientist." At this point, there are five syntactic heads that are required to form a grammatical sentence: 1) the top-level verb; 2) a verb to head the first RC; 3) an NP empty-category to be coindexed with the first RC pronoun "who"; 4) a verb to head the second RC; and 5) an NP empty-category to be coindexed with the second RC pronoun "who." Three new discourse referents have been processed since the prediction of the top-level verb were made ("the student," "the professor," and "the scientist"), two new discourse referents have been processed since the prediction of the verb and empty-category for the first RC ("the professor" and "the scientist"), and one new discourse referent has been processed since the prediction of the verb and empty-category for the second RC ("the scientist"). This quantity of predictions over this many new discourse referents is proposed to be very difficult for the processor to maintain[1].

Empirical evidence for the discourse-based distance metric is provided by Gibson & Warren (in preparation) who used a complexity rating questionnaire to show that doubly nested RC structures are easier to process when a first- or second-person pronoun is in the subject position of the most embedded RC, as in (3), as compared with a similar structure in which an NP introducing a new object into the discourse is in the subject position of the most embedded clause, as in (1) (Bever, 1970; Gibson, 1991; Kac, 1981):

(3) The student who the professor who I collaborated with had advised copied the article.

(4) Isn't it true that example sentences [that people [that you know] produce] are more likely to be accepted? (De Roeck et al., 1982)

The lower complexity of nested structures like (3) and (4) can be accounted for if the memory increment for a predicted category is larger for new discourse referents than for referents that are already part of the current discourse, such as first- or second-person pronouns. (The current discourse always includes a speaker/writer and a hearer/reader.) In particular, if there is no memory cost

---

[1] It is also claimed that the prediction of the top-level verb is cost-free, but this claim is tangential to the issues under consideration here.

increment for predicted heads when a referent that is already part of the current discourse is processed, then only two new discourse referents have been processed since the top-level verb was predicted, only one new discourse referent has been processed since the prediction of the verb and empty-category for the first RC ("the professor"), and no new discourse referents have been processed since the prediction of the verb and empty-category for the second RC. As a result, the memory cost at this point is substantially less than the maximal complexity of processing a sentence like (1). Note that a clause-based proposal does not predict this complexity contrast, because there are three partially processed sentences at the most embedded subject of (3) and (4), just as in (1).

Another example of a construction that violates Kimball's clause-based explanation of interpretive processing is an RC embedded within a sentential complement (SC) of a noun:

(5) SC/RC structure: The possibility that the administrator who the nurse supervised had lost the medical reports bothered the intern.

The acceptability of the SC/RC structure does not follow from the clause-based proposal: There are three partially processed clauses at the point of processing the most embedded subject NP "the nurse" in (5). Thus an SC/RC structure should be just as complex as a doubly nested RC structure, but an SC/RC structure is much easier to (Cowper, 1976; Gibson, 1991; Gibson & Thomas, 1997). This difference is explained by the theory based on incomplete dependencies. Unlike an RC, there is no RC pronoun in an SC, with the consequence that an SC does not require an empty category to occur in the clause. As a result, there are only four syntactic heads required at the most embedded position in (5), and thus its lower complexity is accounted for. Given the shift in recent theory from an interpretive processing mechanism based on clause units to one based on incomplete head-dependency relationships and the number of new discourse referents, it is worthwhile to investigate whether the evidence for clause-based post-interpretive processing might be reinterpreted as well.

Units of post-interpretive processing

Clauses, and not within-clause phrase boundaries, appear to be the units of segmentation during speech perception: Clicks occurring during the auditory presentation of sentences are misheard to occur between clause boundaries but not between phrase boundaries or between individual words (Bever et al., 1969). Controlling for lexical items and serial position, words occurring in recent clauses are recognized more quickly than words occurring in early clauses (Caplan, 1972; Chang, 1982). Early research examining post-interpretive sentence processing, and the capacity of sentence memory, appears to show that the clause is the unit of sentence memory, as well. Several pieces of evidence support this assumption: First, memory declines as a function of the number of clauses in the sentences (Blauberg & Braine, 1974). Second, verbatim recall is highly accurate for the most recent clause heard, but not for earlier clauses (Jarvella, 1971).

However, there are a number of other candidate units of sentence memory: Memory could be a function of the number of words in the sentence, the number of content words (noun phrases (NP), verb phrases (VP), adjectives, etc.), or the number of new discourse referents. Glanzer and Razel (1974) showed that short, familiar sentences (proverbs) could be held as units in short-term memory, but that unfamiliar sentences were recalled less accurately, suggesting that not all clauses can be held as chunks in short-term storage. While it is unlikely that memory will simply be a function of the number of words in the sentence (given the ample experimental evidence (Gershberg & Shimamura, 1995; Larkin & Burns, 1977; Miller, 1956; Miller & Isard, 1964) showing that meaningful word strings are recalled more accurately than unrelated lists), the studies reviewed above do not rule out the other alternatives. The confounds in these studies will be discussed below; because of these confounds, the results do not unambiguously support the assumption that clauses are the units of sentence memory.

One study that has been used to support the idea that clauses are the units of sentence memory was conducted by Blauberg and Braine (1974). Subjects were presented with 30 sentences, six at each length from 2-6 clauses, and then presented with a probe, one of the clauses from the sentence with the subject or object noun missing, and asked to produce the appropriate noun. Recall accuracy declined as a function of the number of clauses in the sentence, with an average of 6/6 probe questions answered correctly for 2-clause sentences, and 3/6 probes answered correctly for 6-clause sentences. While these results show that memory declines as a function of the number of clauses in the sentence, they do not show that the clause is the unit of sentence memory. Sentences with more clauses were longer, and thus it is impossible to determine whether the difficulty with long sentences was due to their greater length, or their greater number of clauses. Determining whether clauses are the units of sentence memory requires a comparison of sentences in which the effects of the number of words and the number of clauses can be separated.

A second study purporting to show that clauses are the units of sentence memory was conducted by Jarvella (1971). He played subjects connected discourse that was periodically interrupted, with subjects instructed to recall as much of the previous material as possible. He examined recall accuracy for the most recently heard two (Experiment 2) or three (Experiment 1) clauses. Overall, he found that the most recent clause was recalled highly accurately, with earlier clauses recalled less accurately, results that appeared to support the idea of clause-based processing. However, there are two problems with the experimental materials and methods that preclude one from interpreting these results as support for clause-based processing. First, the results of Experiment 2 showed that while there was an overall effect of clause position (early vs. recent), when the two clauses were within the same sentence, recall was still accurate for the earlier clause, (early clause = .84, recent clause = .97), whereas when they were in different sentences, recall was much more accurate in the most recent clause (.95) than in the earlier clause (.63). This result provides more support for the idea that sentences are the units of memory, rather than clauses. But problems with the materials render even this interpretation problematic: When the recent and early clauses were in different sentences, the sentence containing the early clause was more complex than when the recent and early clauses were in the same sentence. Sample sentences from Jarvella (1971) illustrate this problem:

(1) Early clause in same sentence: He and the others were labeled as Communists. McDonald and his top advisors hoped this would keep Rarick off the ballot.

(2) Early clause in previous sentence: That he could be intimidated was what McDonald and his top advisors hoped. This would keep Rarick off the ballot.

While the words in the final two clauses ([MacDonald and his top advisors hoped] [this would keep Rarick off the ballot]) are identical in the two conditions, the sentence structure in (2) is much more complex. The first sentence in (2) contains the cleft sentential subject "that he could be intimidated," which is much harder to understand than the simpler subject-verb-object sequence in (1) (Frazier & Rayner, 1988; Gibson, in press). This factor would render the sentences in (2) much harder to recall, especially to recall verbatim, as was required in this study. Measuring verbatim recall is the second weakness of this experiment: Requiring subjects to recall sentences verbatim focuses on the surface structure of sentences, rather than the conceptual or propositional content. Potter and colleagues have provided evidence that verbatim recall, relying on a briefly held phonological record, can be dissociated from memory for the conceptual content of sentences, memory that retains an abstract representation of propositional content without retaining the exact lexical or syntactic form of the sentence (Lombardi & Potter, 1992; Potter, 1993; Potter & Lombardi, 1990; Potter, Moryadas, Abrams, & Noel, 1993; Potter, Valian, & Faulconer, 1977). Even if the experiments in Jarvella (1971) had succeeded in showing that clauses are the units of verbatim recall, this finding could not be used as evidence that clauses are the units of sentence memory when sentence content is probed, rather than surface form and lexical items.

Although the evidence that the clause is the unit of segmentation during speech perception is convincing (Bever et al., 1960; Caplan, 1972; Chang, 1982), neither Blauberg & Braine (1974) or Jarvella (1971) provide conclusive evidence that the clause is the unit of sentence memory. Given the recent evidence showing that discourse referents are the units of interpretive processing (Gibson, 1998), I considered it worthwhile to conduct an experiment to determine whether the unit of post-interpretive processing is (a) the clause or (b) the discourse referent.

The Role of Working Memory in Memory for Sentences

Working memory and understanding language

The most influential model of working memory is the multicomponent (MC) model proposed by Baddeley and colleagues (Baddeley, 1996; Baddeley, 1983;

Baddeley & Hitch, 1994), which has been extended and modified by other researchers (Lehto, 1996; Martin & Romani, 1994; Smith & Jonides, 1997). In MC models, the verbal part of working memory consists of at least two components: The short-term store (STS), used for storing and rehearsing verbal information using a phonological code[2] (Awh et al., 1996; Baddeley, 1996; Basso et al., 1982; Fiez et al., 1996; Paulesu et al., 1993; Vallar, Betta, & Silveri, 1997); and the central executive (CE), used for allocating attention, planning, inhibiting nonrelevant responses, and coordinating resources demanded by concurrent tasks (Baddeley, 1996; D'Esposito et al., 1995; Lehto, 1996). Answering questions about sentences could depend primarily on the STS, the CE, or both. One goal of the current experiments is to determine whether the STS and the CE make independent contributions to sentence memory capacity.

Gathercole & Baddeley have hypothesized that the STS and the CE make dissociable contributions to language understanding (Gathercole & Baddeley, 1993). For sentence processing, the STS is used to maintain a phonological record of sentences just heard or read, a record that can be consulted off-line during post-interpretive processing. Such a record would be particularly useful for sentences that are initially understood incorrectly and must be reanalyzed (i.e., The cotton clothing is made of is grown in Mississippi), or that contain a long list of items to be remembered (i.e., Please go to the store and buy bread, milk, eggs, cheese, oranges and spinach). Evidence for the role of the STS in post-

---

[2] In Baddeley's model, passive verbal storage includes two components: the phonolgical store, used for holding information, and the articulatory loop, used to rehearse information using a speech-based code (Baddeley, Vallar, & Wilson, 1987; Baddeley, 1983); . Such a fine-grained

interpretive processing includes the finding that subjects are impaired at comprehending long, complex sentences when they have to concurrently articulate irrelevant words (articulatory suppression) (Baddeley, Eldridge, & Lewis, 1981).

An early idea about the role of STS in language comprehension held that sentence comprehension requires an ordered, verbatim representation of the words just heard. However, this idea was contradicted by the discovery that patients with a severe deficit in the STS (digit or word spans of one to four items) are relatively unimpaired in understanding language (Baddeley et al., 1987; Basso et al., 1982; Belleville et al., 1992; Martin, 1987; Martin, 1993; McCarthy & Warrington, 1987a; McCarthy & Warrington, 1987b; Saffran & Marin, 1975; Vallar et al., 1997). When tested in detail, such patients show impairment only in understanding very long or complex sentences (see Caplan & Waters, 1990, for a review). Other results that questioned the relevance of a verbatim representation for sentence comprehension come from Potter and colleagues, who found that people quickly form conceptual representations of sentence meaning while losing information about the exact words and syntactic structure (Lombardi & Potter, 1992; Potter, 1993; Potter & Lombardi, 1990; Potter et al., 1993; Potter et al., 1977). Such evidence indicates that the STS is not crucial for first-pass comprehension, although it may be useful for higher-level linguistic interpretations that lag behind on-line comprehension processes.

---

analysis is not relevant to the hypotheses under investigation here, so for the sake of simplicity we are including both components in the STS.

According to Gathercole & Baddeley, the CE is used for syntactic and semantic processing, and for storing the intermediate and final products of such processing during performance of post-interpretive tasks (i.e., answering questions, verifying the truth of statements, reasoning from given propositions). The CE has also been claimed to be important for integrating new propositions with representation of a text and maintaining the predicate-argument structure of propositions (Caplan & Waters, 1998; Gathercole & Baddeley, 1993; Kintsch & van Dijk, 1978). Evidence for the role of the CE in sentence memory includes the fact that tasks like Daneman & Carpenter's (1980) Reading Span, which require the coordination of storage and processing (a CE function), correlate more highly with measures of reading comprehension (i.e., verbal SAT, answering factual questions about a passage, or understanding pronoun referents) than do measures of STS alone (i.e., digit or word span). Another source of evidence about the role of the CE in sentence memory comes from examining the performance of patients with impaired CE function. Patients with Alzheimer's disease (AD) have severe impairments in CE functions (Baddeley et al., 1991; Baddeley, Logie, Bressi, Sala, & Spinnler, 1986). Waters and colleagues (Waters & Caplan, 1997; Waters, Caplan, & Rochon, 1995) have presented evidence that patients with AD are impaired in post-interpretive processing of sentences with more than one proposition, and are more disrupted than control subjects under dual-task conditions, which require the coordinating function of the CE. These results suggest that the CE may be crucial for normal post-interpretive sentence processing.

Measuring working memory

Any researcher attempting to measure working memory capacity faces a serious challenge in deciding what test to use. A review of the working memory literature shows that several very different tasks are commonly used to assess working memory (Baddeley et al., 1985; Braver et al., 1997; Case et al., 1982; Daneman & Carpenter, 1980; Daneman & Merikle, 1996; Dobbs & Rule, 1989; Engle et al., 1992; Klapp et al., 1983; LaPointe & Engle, 1990; Petrides et al., 1993b; Salthouse, 1990; Turner & Engle, 1989; Waters & Caplan, 1996b). Several recent studies, however, have suggested that scores on commonly used working memory tests may not in fact be highly correlated, and that the different tests may be differentially sensitive to processes such as storage, response time, rapid stimulus manipulation, and other CE functions (Dobbs & Rule, 1989; Lehto, 1996; Waters & Caplan, 1996b). The experiment reported in Chapter 2 compared performance on several different tests of working memory, and found that they were not highly intercorrelated, and that scores on different tests were predicted by measures of different component variables (short-term storage and processing speed). These results suggest that different working memory tests are sensitive to different components of working memory.

Determining whether the STS and the CE make independent contributions to sentence memory requires independent measures of the two components. For the STS, such a measure is straightforward: STS capacity is measured using a task like word span or digit span that requires subjects to recall items exactly as presented, with no manipulation or computation required. However, measuring the CE provides more of a challenge: There are no widely accepted and uncontroversial measures of working memory itself (Baddeley et al., 1985; Waters & Caplan, 1996b), and most of the commonly used working memory tests require subjects to store as many items as possible while performing computations or manipulations. For example, backward digit span is considered to be a test of working memory, rather than simply a measure of STS capacity, in that it requires subjects perform a manipulation, order reversal, on the contents of storage. But the measure of capacity is the number of digits that can be recalled in reverse order, which will probably depend on the capacity of the STS. In Daneman & Carpenter's (1980) Reading Span test, subjects are required to store as many sentence-final words as possible while concurrently processing sentences. Performance on such tasks depends not only on how efficiently a subject can process or manipulate stimuli (CE functions), but also on how many items a subject can store (capacity of the STS).

There are a few working memory tests that appear to be relatively pure measures of the CE independent of the capacity of the STS. These tests require subjects to monitor and manipulate stimuli, or update the contents of storage, but do not require subjects to store a large number of items: Any storage requirements of such tasks is well below the limits of the STS. For example, in the N-back test, subjects are asked to monitor a string of stimuli, and to respond when a target is presented, with a target defined as an item that is the same as one that occurred n items ago, or "N-back." This task has been used extensively in neuroimaging studies, and has been shown to activate areas in prefrontal cortex thought to be the neural substrate of the CE component of working memory (Awh et al., 1996; Braver et al., 1997; D'Esposito et al., 1995; Schumacher et al., 1996; Smith & Jonides, 1997). In the 2-back and 3-back tests most commonly used, the number of items that must be stored is well below subject's STS capacity, but requires subjects to maintain an ordered representation of the last N items presented, to compare incoming items to the appropriate stored item, and to update the stored items on each trial. The results in Chapter 2 demonstrated that N-back does not correlate with STS capacity (measured by digit span or word span) whereas two other working memory tests, backward digit span and reading span, do correlate with STS. These results suggest that N-back is a relatively pure measure of CE function. I addressed the question of the relative contribution of the STS and the CE to sentence memory capacity by examining (a) the correlation between Word Span (a measure of STS) and sentence memory, (b) the correlation between N-back (a measure of the CE) and sentence memory, and (c) a multiple regression of Word Span and N-back on sentence memory, to determine whether each test accounts for independent variance in the sentence memory score.

The relation between working memory and language comprehension or other cognitive functions has been assessed using two different analytic techniques: correlational studies and the individual differences approach. In correlational studies (Babcock & Salthouse, 1990; Case et al., 1982; Daneman & Carpenter, 1980; Daneman & Carpenter, 1983; Daneman & Merikle, 1996; Engle et al., 1992; LaPointe & Engle, 1990; Turner & Engle, 1989; Waters & Caplan, 1996b), subjects are given a variety of tests of working memory and other cognitive functions, and correlations, multiple regressions, and factor analyses are used in order to determine what variables are related to the cognitive function of interest. These practices allow researchers to determine relations among large numbers of variables, and the extent to which different tests contribute common and unique variance to the measure of interest (Engle et al., 1992). One limitation of correlational approaches is that multiple comparisons require large numbers of subjects in order to be reliable. Another problem in the literature (although not inherent in the approach) is that with large numbers of subjects, statistically significant correlations may account for only a small amount of the variance on a given test. For example, with a large number of subjects, an $\underline{r}$ of .25 may be statistically significant at the $\underline{p} < .05$ level, but would explain less than 10% of the variance. Such a small correlation would not constitute a sufficient explanation of the relation between the correlated variables.

The second approach to studying the relation between working memory and language is the individual differences approach (Just & Carpenter, 1992; Just et al., 1996a; King & Just, 1991; King & Kutas, 1995; MacDonald, Just, & Carpenter, 1992; Miyake, Carpenter, & Just, 1994a; Miyake, Carpenter, & Just, 1995; Miyake, Just, & Carpenter, 1994b). In this method, subjects are given a test designed to measure working memory capacity, such as Reading Span, and then divided on the basis of their scores into three groups, a high-, medium-, and low-span group. Usually, the medium-span group is omitted from further analyses, and the high- and low-span groups are compared, using ANOVAs, on another measure of interest, such as reading speed or sentence comprehension. The groups are treated as though they are independent and homogeneous, and are compared to see whether they perform differently on the secondary task (Engle et al., 1992). This approach, however, has some significant flaws: First, leaving out the subjects in the middle of the sample ignores a large amount of data, including information about the variability of the sample. Second, this approach may in fact lead to an overestimation of the relation between variables. Selecting only extreme groups eliminates those subjects whose scores would be near the mean of the sample, leaving subjects whose scores have larger deviations from the mean. As a result, the correlation coefficient is likely to be larger with extreme groups (for a discussion see McCall, 1998, pp. 168-169). Third, the choice of a cutoff point seems arbitrary if inspection of a scatterplot does not suggest any natural grouping of the data. For these reasons, the correlational approach is superior unless scatterplots show a natural grouping of the data.

Sentence memory without working memory

MacDonald and Christiansen (1998) have recently presented an alternative to the currently dominant view of the relation between working memory and language processing. They claim that there is no linguistic working memory capacity separate from linguistic representations and processes. In this view, measures of language processing and measures of linguistic working memory are simply different measures of language processing skill. Individual differences supposedly due to differences in working memory capacity are due to differences in skill and experience with language.

MacDonald & Christiansen point out that "the fact that subjects are tested on tasks that are called 'working memory tasks' does not entail that the construct of a working memory separate from processing is a valid one" (p.3). While this statement is certainly true, and MacDonald & Christiansen's hypothesis has the appeal of offering an alternative to relying on differences in a poorly defined and measured working memory capacity, this view does not provide a convincing alternative that explains the existing data. First, it is not clear that it is actually an alternative: the explanation for individual differences (differences in language processing skill) translates easily into working memory models such as those of Just and Carpenter (1992), or Salthouse (1990), which view working memory capacity as the interaction of storage capacity and processing efficiency. In these models, individual differences on working memory tests could be due to differences in storage capacity, processing efficiency, or both. Thus MacDonald & Christiansen's alternative could be seen as a case of differences in a specific kind of processing efficiency (reading skill) explaining individual differences in language processing tasks and linguistic working memory tasks.

MacDonald & Christiansen's account, however, addresses only the relation between language processing and linguistic measures of working memory, such as Daneman and Carpenter's (1980) Reading Span, or the auditory analog, Listening Span, both of which involve reading and remembering the final words of sentences. The fact that these tasks require reading or listening to sentences makes it plausible to suppose that any individual differences observed may be due to differences in reading skill. But this explanation would not account for the correlations that have been observed between reading comprehension and nonlinguistic tasks of working memory (Daneman & Merikle, 1996; Turner & Engle, 1989; Engle et al., 1992). A skill-via-experience account, in which better readers do better in reading comprehension and in linguistic working memory tasks, offers no explanation for correlations among linguistic and non-linguistic working memory tasks, and no explanation for correlations between working memory, as measured by these tasks, and sentence memory.

Experiment

The experiment consisted of two parts: Part 1 tested subjects' sentence memory capacity. Over headphones, subjects heard sentences of different lengths, from two to five clauses. Sentences were semantically unconstrained: Any agent could plausibly perform any action in the sentence. Immediately following the sentence, subjects heard a question, probing their memory for one of the clauses in the sentence. In order to ensure that subjects paid attention to all parts of the sentence, two types of probes were presented: questions probing memory for either the subject of the clause (Agent questions) or the main verb (Action questions). For all clauses in the sentences except the final clause, there were an equal number of probes at each serial position. The final clause of the sentence was never probed.

In order to address the question of the unit of memory, three different sentence types were assessed: Sentences with relative clauses (RC), sentential complements (SC) , and relative clause with double objects (DO) (two NPs or 1 NP and 1 PP) . DO sentences with the same number of clauses were longer (containing one additional new discourse referent per clause) than RC and SC sentences. In fact, DO sentences at a given length (n) contained the same number of new discourse referents as RC and SC sentences at length n+1. Thus, if memory is a function of the number of clauses, there should be no difference between DO sentences and the other two sentence types. However, if memory is a function of the number of new discourse referents in the sentence, then accuracy on DO sentences should be worse than the other two sentence types: DO sentences of length n should be recalled as poorly as the other two sentence types at length n+1.

Part 2 tested subjects' short-term storage and working memory capacity. Several different working memory measures were used, in order to explore the relation among them and replicate the results in Chapter 2. The tests are described in the Methods section, and zero-order correlations are reported in the Appendix. However, in order to facilitate interpretation of the results and to reduce the number of comparisons to a statistically permissible level, the discussion focuses on a subset of those tests: Word Span, Backward Digit Span, N-back, and a Complex Span measure derived by forming a composite score from three different tests modeled on Daneman & Carpenter's (1980) Reading Span test: Reading Span, Math Span, and Category Span. The tests are described in detail in the Methods section; the composite score was used in order to obtain a measure of simultaneous storage and processing capacity that was independent of the processing task. In order to form a composite measure, it was necessary to modify Daneman & Carpenter's original procedure so that the task demands of the three tests were identical: The only difference between them was the background processing task and the item to recall. Combining several tasks into a composite measure rules out the possibility that any observed relation between the variables might be due to the materials used in this study, or to the fact that Reading Span and sentence memory may both be sensitive to a third factor, subject's reading skill (MacDonald & Christiansen, 1998). Controlling for these alternative explanations is especially important given that there are so many versions of the original Reading Span measure, and very few attempts have been made to validate and compare different versions (Baddeley et al., 1985; Daneman & Merikle, 1996; Turner & Engle, 1989; Waters & Caplan, 1996b).

Method

Part 1: Sentence Memory

Subjects

Thirty MIT students served as subjects. All subjects were native English speakers. Subjects were paid $8 for their participation.

Materials

Each subject heard 121 sentences of four different lengths, 2-clause, 3-clause, 4-clause, and 5-clause. Three types of sentences were used: RC, SC, and DO. The Question Type (Action vs. Agent) and Probe Location (1st, 2nd, 3rd, or 4th clause) variables were balanced across the sentence types, such that equal numbers of the two question types and locations were probed for each sentence type. Order was randomized, then the same set of sentences were presented to each subject. A sample 3-clause sentence of each type is presented in Table 1. All clauses, except the final clause, were semantically unconstrained: Sentences were constructed using a program that randomly assigned subjects with verbs for all but the final clause in the sentence. The final clause (which was never probed) was constructed to be semantically plausible to provide some conceptual closure to the sentence. Except for the final clause, each serial position was probed equally often. For RC and DO sentences, probes consisted of two question types, in order to optimize comprehension and attention to the entire sentence: Agent questions (Who lectured someone?) and Action questions (What did the barber do?). Because SC sentences did not have a single NP object, only Agent questions were used.

Table 8. Sample 3-Clause Sentences

Relative Clause (RC)	The barber lectured the sailor who hit the singer who worked in the jazz club.
Sentential Complement (SC)	The violinist insisted that the immigrant doubted that the chef had trained in Paris.
Double Object (DO)	The psychologist showed the document to the criminal who sent a gift to the editor who was compiling an anthology.

Procedure

Sentences were recorded onto a Macintosh Quadra 640 computer using Sound Designer II software, and were played back to the subject over headphones. Sentences were read naturally. Immediately following sentence presentation, subjects heard a question about the sentence, which they answered aloud. Subjects pressed the spacebar when they were ready for the next sentence. The experimenter marked the accuracy of the subject's response on a scoresheet. Agent questions were scored 1 if they were answered correctly and 0 if they were answered incorrectly. Action questions were scored 1 if both parts of the answer were correct (the verb and the object), .5 if one part was correct, and 0 if both parts were incorrect or omitted. The experimental session began with 6 practice sentences, followed by the 121 test sentences. The session lasted approximately 30 minutes.

Part 2: Working Memory

Subjects

Twenty-six subjects who had participated in Part 1 returned on a second day to participate in Part 2.

Materials

Short-Term Storage Tests

Forward Digit Span (fDS). Subjects heard strings of digits presented at the rate of 1 digit per sec, and then recalled them in order. Span was defined as the longest string of digits a subject could repeat correctly, in order, on one of two trials.

Word Span (WS). Subjects heard lists of words (1-syllable concrete nouns) presented at the rate of 1 word per sec, and then recalled them in order. Span was defined as the longest string of words a subject could repeat correctly in order on one of two trials.

Working Memory Tests

Backward Digit Span (bDS). Subjects heard strings of digits presented at the rate of 1 digit per sec, and then recalled them in reverse order. Span was defined as the longest string of digits a subject correctly repeated in reverse order on at least one of two trials.

N-back. Subjects saw words (4-letter abstract nouns) presented one at a time on the computer screen at the rate of one word every 3 sec (2500 ms word presentation, 500 ms interstimulus interval). They responded with a button press whenever they saw a target. A target was defined as a word that was the same as the word presented N ago, or "N-back." Subjects were first presented with 2-back targets; if they reached criterion (70% correct) they were presented with 3-back targets, then 4-back, and then 5-back. There were 70 - 80 trials at each set size, with 10 correct targets (hits) per set. The score for each set size was computed by subtracting the number of false alarms from the number of hits to correct for guessing; then scores for all levels completed were combined to reach a composite N-back score. The equation for combination was as follows: 1 + ((2-back, % correct) + (3-back, % correct) + (4-back, % correct) + (5-back, % correct) X 100).

Reading Span. Subjects viewed sets of short declarative sentences (5-10 words, mean 7.3) on the computer screen, and read them aloud. Next, subjects viewed simple questions (probing either the subject or the main verb) and answered them aloud. After two sentence-question sets, subjects were prompted to recall the final word of both sentences. Subjects were first presented with five trials at set-size two; in order to advance to larger set sizes (three to six), they had to recall all the words correctly on three of the five trials. Span was defined as the largest set size at which subjects recalled all of the words correctly on four of the five trials; with an additional .2 added for each trial they recalled correctly at the next set size.

Math Span (MS) – The procedure was the same as for Reading Span, except that subjects saw a simple addition problem, and reported the sum aloud. After two such trials, subjects recalled the second digit of each of the two problems aloud. If the subject recalled the two digits correctly on three of five trials at set size two, the set size was increased to three. The largest possible set size was six. Scoring was the same as for Reading Span.

Category Span. The procedure was the same as for Reading Span, except that subjects read a list of four nouns, three of which belonged to a common category (i.e., animals, foods or colors). The fourth word did not match the category. On each trial, subjects reported the category name aloud. After two such trials, subjects recalled the mismatch word for the two lists. If the subject recalled the two words correctly on three of five trials at set size two, the set size was increased to three. The largest possible set size was six. Scoring was the same as for Reading Span.

Counting Span (CountSp) – The procedure was the same as for Reading Span, except that subjects saw sets of yellow and blue dots on the screen, counted the yellow dots, and reported the number aloud. After two such trials, subjects said aloud the number of yellow dots that they had counted on each of the two screens. If the subject recalled the two numbers correctly on three of five trials at set size two, the set size was increased to three. The largest possible set size was six. Scoring was the same as for Reading Span. Notice that this task differs from the three above tasks: Subjects report the same thing after the set of trials (the number of yellow dots on each trial) as they report aloud following each trial. In the previous tests, the item to recall following the set of trials was different from the item reported on each trial.

Procedure

Subjects were tested in the following order: forward Digit Span, backward Digit Span, Word Span, N-back. Testing order was held constant for all subjects because we planned to use correlations between scores and thus wanted factors such as practice and fatigue to be held constant. Then, the four complex span measures were presented in pseudorandom order, with the condition that the two tests requiring word recall (Reading Span and Category Span) and the two tests requiring number recall (Math Span and Counting Span) never occurred consecutively. Testing order for these for the complex span measures was randomized because we planned to combine scores into a composite that would equally reflect the contribution of each score; thus fatigue and practice effects should be distributed equally across the four tests. The experimental session lasted approximately 1 hour.

Data Analysis

For RC and DO clauses, ANOVAs were performed at each clause length to determine whether there was an effect of probe question type (Action vs. Agent). There were no significant differences between question types at any clause length, so for all further analyses the two question types were combined.

For the purpose of correlations, the mean score on the 3-clause sentences was used, because these scores showed the largest individual differences. Scores at the other sentence lengths might have restricted ranges due to ceiling and floor effects. Correlations for the mean overall sentence memory score and the mean score on 3-clause sentences are reported in the Appendix. For the tests discussed below, there is no difference in results for the 3-clause score vs. the overall score.

The composite Complex Span score was formed by first computing a z score for each subject on Reading Span, Math Span, and Category Span, then computing the linear combination of the z scores on the three tests. The linear combination of z scores was used so that each test contributed equally to the Complex Span score.

Results

Figure 1 plots sentence memory as a function of number of clauses for the three sentence types. Error bars show 95% confidence intervals. The means, standard deviations, and ranges are shown in Table 2. Large individual differences were observed, especially for 3-clause sentences. Significant main effects were found for number of clauses, F(3, 87) = 322.99, $p < .001$, and sentence type, F(2,58) = 4.34, $p < .02$; the interaction was also significant, F(6,174) = 2.62, $p < .03$.

Table 2. Percent Correct for Each Sentence Type and Length: Means, (Standard Deviations), and Ranges

	2 Clauses	3 Clauses	4 Clauses	5 Clauses
Relative Clause (RC)	97 (5) 82 – 100	64 (17) 31 – 94	36 (17) 0 – 79	24 (13) 3 – 53
Sentential Complement (SC)	93 (12) 67 – 100	64 (26) 17 – 100	26 (19) 0 – 67	27 (12) 8 – 50
Double Object (DO)	92 (13) 56 – 100	56 (24) 13 – 100	32 (12) 8 – 58	23 (9) 6 – 41

The primary question of interest was whether accuracy on DO sentences, in which each clause contained one additional discourse referent, would be significantly worse than accuracy on the other two types. Figure 2 shows that error bars overlapped at all sentence lengths, suggesting that there was no reliable difference between sentence types. Planned comparisons showed marginally significant differences between the sentence types at the 2-clause (F(2,58) = 2.72, $p < .07$) and 3-clause (F(2,58) = 2.8, $p < .07$) lengths, but no

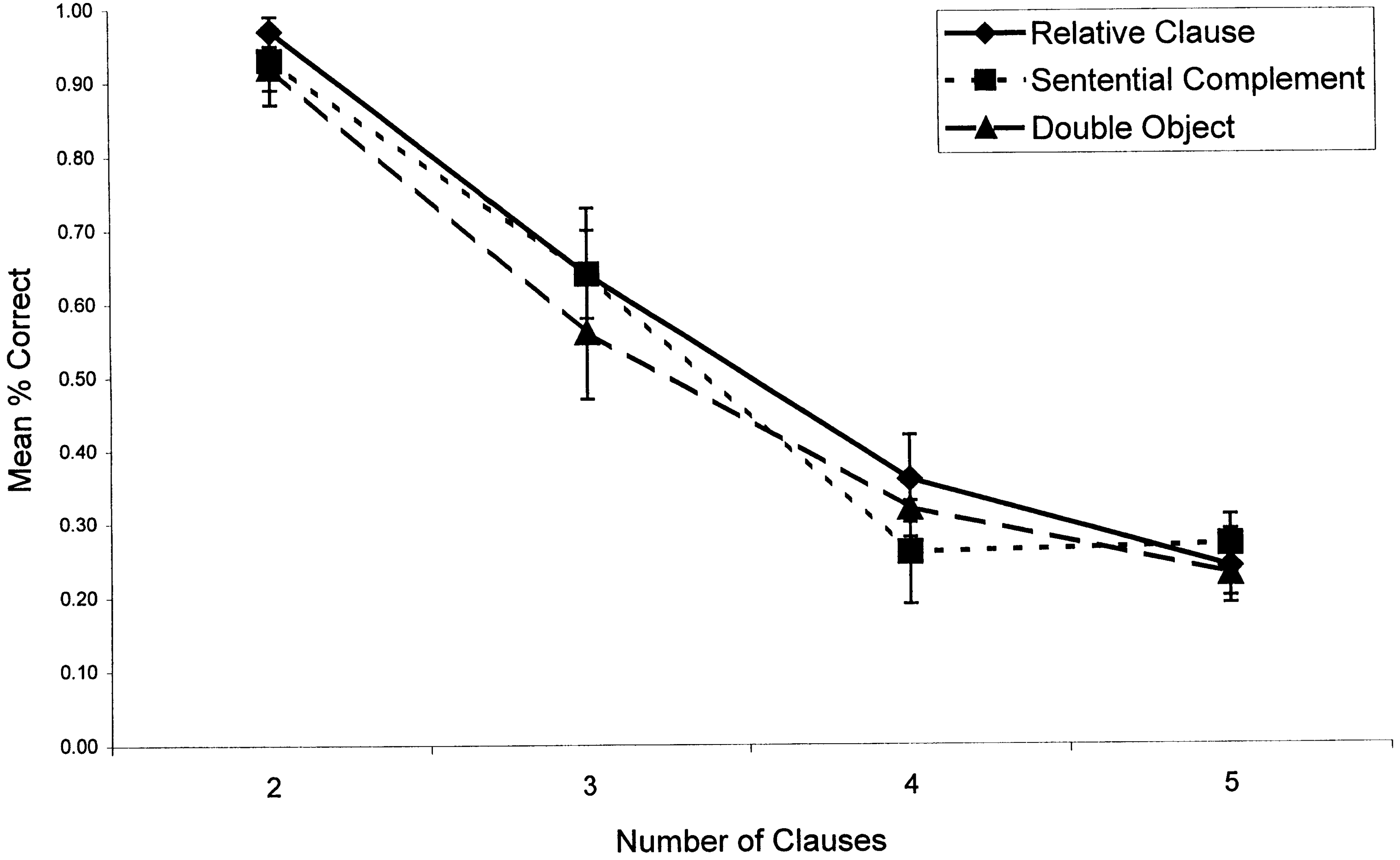

Figure 1. Accuracy of sentence memory as a function of sentence type and number of clauses

consistent finding of significantly impaired performance on DO sentences relative to the other sentence types. Instead, accuracy declined as a function of the number of clauses for all three sentence types. The difference between 4- and 5-clause sentences was significant overall, F(1,29) = 10.5, $p < .003$. This difference reflects the fact that accuracy on RC and DO sentences continued to decline as sentence length increased (RC: 4-clause, 36%; 5-clause, 24%; DO: 4-clause, 32%; 5-clause, 23%) while accuracy on SC sentences did not (4-clause, 26%; 5-clause, 27%).

A table showing zero-order correlations between all working memory measures, the composite Complex Span score, the overall sentence memory score, and the mean sentence memory score for 3-clause sentences is included in the Appendix. I will discuss a subset of those scores here. Table 3 shows correlations among the working memory measures (Backward Digit Span, N-back, and Complex Span) and the STS measure (Word Span). All correlations were significant except the correlation between N-back and Word Span, and N-back and Complex Span.

Table 3. Correlations among Working Memory and Short-Term Storage Measures

	Back Digit Span	N-back	Complex Span
Word Span	.44*	.21	.57**
Back Digit Span		.43*	.46*
N-back			-.09

* $p < .05$
** $p < .01$

Table 4 shows correlations between working memory and STS and sentence memory capacity. All correlations were significant. In order to determine the relative contribution of these variables to explaining individual differences in working memory capacity, I entered the four variables into a multiple regression equation (Table 5): The linear combination of Backward Digit Span, Word Span, N-back and Complex Span significantly predicted sentence memory ($F(4,21) = 5.57$, $p < .003$), explaining 51% of the variance in the sentence memory score. However, the only two predictors that contributed to the relation were N-back, which uniquely accounted for 16% of the variance, and Complex Span, which accounted for an additional 12%. In spite of correlating significantly with sentence memory, Backward Digit Span and Word Span did not account for any additional unique variance. When N-back and Complex Span were entered alone as predictors, the model still accounted for 51% of the variance in sentence memory, and both predictors explained 28% of the unique variance in sentence memory capacity (Table 6).

Table 4. Correlations Between Short-Term Store and Working Memory Measures and Sentence Memory Capacity

STS / WM Test	Correlation with Sentence Memory
Word Span	.44*
Back Digit Span	.50**
N-back	.48**
Complex Span	.48**

* $p < .05$
** $p < .01$

Table 5. Multiple Regression of Working Memory and STS Variables: Percentage of Variance in Sentence Memory and Significant p Values

R^2 for model = .51

STS / WM Test	% Variance*	p
N-back	16	.01
Complex Span	12	.03
Word Span	<1	n.s.
bDS	<1	n.s.

*Squared semi-partial correlations

Table 6. Multiple Regression of N-back and Complex Span: Percentage of Variance in Sentence Memory and Significant p Values

R^2 for model = .51

STS / WM Test	% Variance*	p
N-back	28	.001
Complex Span	28	.001

*Squared semi-partial correlations

In order to determine whether the different components of working memory each contribute to sentence memory, I entered the two tests hypothesized to measure the components independently: N-back as a measure of the CE, and Word Span as a measure of the STS (Table 7): The linear combination of N-back and Word Span significantly predicted sentence memory ($F(2,23) = 6.29$, $p < .007$), explaining 35% of the variance in the sentence memory score. Each predictor uniquely accounted for a significant portion of the variance: N-back for 16% and Word Span for 12%.

Table 7. Multiple Regression of Individual Components of Working Memory: Percentage of Variance in Sentence Memory and Significant p Values

R^2 for model = .35

Component Test	% Variance*	p
N-back (CE)	16	.03
Word Span (STS)	12	.05

*Squared semi-partial correlations

Discussion

The first question I addressed was the unit of sentence memory. I tested two alternative hypotheses: (a) that sentence memory would be a function of the number of clauses and (b) that sentence memory would be a function of the number of discourse referents. For interpretive processing, discourse referents appear to be the units of processing (Gibson, 1998); however, the current results show that for sentence memory (one form of post-interpretive processing) the clause, rather than the discourse referent, is the unit of memory. This finding not only confirms, in a rigorous way, the commonly held belief in that the unit of sentence memory is the clause, but also is relevant to the debate, discussed in Caplan & Waters (1998), about whether interpretive and post-interpretive processing use the same memory resources. The fact that the two types of processing use different units suggests that Caplan and Water's suggestion, that the two types of processing are distinct, is correct.

The second question concerned the relation between sentence memory capacity and working memory, and whether the two hypothesized components of verbal working memory (the STS and the CE) both contribute to sentence memory capacity. I found that over half the variance in sentence memory capacity could be explained by a combination of two working memory tests, the N-back, which I have hypothesized measures one aspect of CE function, and the Complex Span measure, a composite of three tests modeled on Daneman & Carpenter's (1980) original Reading Span measure. Because Complex Span is a combination of three tests with similar task demands (simultaneously storing and processing information and switching attention between the two subtasks) but with different processing requirements (reading sentences, performing addition problems, categorizing words), the predictive power of Complex Span reflects the relation between sentence memory and whatever aspect of working memory that Complex Span measures. This finding contradicts MacDonald & Christiansen's (1998) claim that correlations between linguistic working memory measures and sentence comprehension measures are due to the fact that both are sensitive to subject's reading ability.

These results are also relevant to MacDonald & Christiansen's (1998) attempt to abolish the working memory construct. Their alternative skill-via-experience account (better readers are better at both linguistic working memory and sentence comprehension tasks) does not account for the correlations observed in this study. N-back, Math Span, and Category Span, tests that did not involve reading sentences, were correlated with sentence memory. In fact, the zero-order correlations (shown in the Appendix) between Math Span and Category Span and sentence memory were higher than between Reading Span and sentence memory. MacDonald & Christiansen's only possible explanation for such correlations is that all these tests are sensitive to "the accuracy of phonological representations," which they claim is a biological factor underlying individual differences (along with reading skill, an experiential factor). In fact, their explanation for individual differences on Reading Span and Listening Span is that these tests reflect differences in phonological processing ability, rather than differences in working memory. However, once again this account shifts the burden of explaining individual differences from the working memory construct to the idea that "maintaining a set of unrelated words requires substantial activation of phonological representations" (p. 14), as does sentence comprehension. They claim that Reading Span measures "the ability to comprehend sentences in the face of competing phonological activation from a series of words that are being prepared for articulation, and to maintain phonological activation for the words in the face of competing demands from sentence processing" (p. 15). Maintaining phonological activation of words is another way of describing the storage functions of working memory. Thus, MacDonald and Christiansen have not, in fact, presented an adequate alternative to the idea of variance in working memory capacity as the source of individual differences in understanding language.

Although the dominant view of working memory and sentence comprehension does not provide a complete and satisfying account for individual differences observed on working memory tests, or the relation between scores on working memory tests and language processing, at this time there does not appear to be a viable alternative view. The current finding, that there is a correlation between working memory tests and sentence memory, suggests that they are both calling on resources that are central, in that the same resources are used for a variety of complex cognitive tasks. Much more research will be required to (a) provide better definitions of working memory, especially those functions subserved by the CE in Baddeley's (1986) model, and (b) determine what factors (cognitive or neurological) account for the individual differences observed on different working memory and sentence memory tests.

Appendix

Zero-Order Correlations for All Tests Used in Chapter 3

	fDS	WS	bDS	N-back	Comp Span	Read Span	Math Span	Cat Span	Count Span	SentMem (all)	SentMem (3 CI)
FDS		0.29	0.71***	0.2	0.40*	0.25	0.34	0.39*	0.22	0.29	0.31
WS	0.29		0.44*	0.21	0.57**	0.52**	0.28	0.58**	0.12	0.43*	0.44*
BDS	0.71***	0.44*		0.43*	0.46*	0.32	0.44*	0.37	0.19	0.48**	0.50**
N-back	0.20	0.21	0.43*		-0.09	-0.15	0.1	-0.17	0.24	0.49**	0.48**
Comp Span	0.40*	0.57**	0.46*	-0.09		0.88***	0.74***	0.81***	0.28	0.44*	0.48**
Read Span	0.25	0.52**	0.32	-0.15	0.88***		0.48**	0.65**	0.23	0.31	0.33
Math Span	0.34	0.28	0.44*	0.10	0.74***	0.48**		0.33	0.28	0.37	0.49**
Cat Span	0.39*	0.58**	0.37	-0.17	0.81***	0.65**	0.33		0.17	0.39*	0.36
Count Span	0.22	0.12	0.19	0.24	0.28	0.23	0.28	0.17		0.16	0.31
SentMem (all)	0.29	0.43*	0.48**	0.49**	0.44*	0.31	0.37	0.39*	0.16		0.92***
SentMem (3 CI)	0.31	0.44*	0.50**	0.48**	0.48**	0.33	0.49**	0.36	0.31	0.92***	

*	$p < .05$
**	$p < .01$
***	$p < .001$

Chapter 4

Conclusions

If little else, the brain is an educational toy. While it may be a frustrating plaything – one whose finer points recede just when you think you are mastering them – it is nonetheless perpetually fascinating, frequently surprising, occasionally rewarding, and it comes already assembled; you don't have to put it together on Christmas morning.

-Tom Robbins, *Even Cowgirls Get the Blues*

I. Modeling and Measuring Working Memory

The experiment reported in Chapter 2 showed that, although the all of the tests compared fit the definition of working memory, scores on different tests were not correlated, and were sensitive to different underlying factors. The fact that different "working memory" tests were uncorrelated suggests that the operational definition of working memory is inadequate: The definition is imprecise enough to include unrelated tests with different task demands. Thus, one prerequisite for progress in understanding working memory is a more precise, restrictive definition of the function. An improved definition of working memory would have to directly address the grab-bag, homuncular nature of the CE component. Many executive functions contribute to working memory: planning; managing goals, coordinating component processes; monitoring automatic processes; controlling effortful processes; marshaling, allocating, and switching attention; and inhibiting prepotent responses (Baddeley, 1996; Lehto, 1996; Petrides, 1994, 1995b). These functions, however, are necessary for the performance of most complex cognitive tasks, not just tasks requiring the on-line maintenance and manipulation of information. One suggestion for making the definition of working memory more specific, and thus more useful, is to restrict it to the on-line maintenance and manipulation of currently relevant stimuli, and to remove those non-mnemonic functions from under the working memory umbrella. Baddeley (1996) has suggested a similar approach: If researchers are able to separate and understand subcomponents of executive control, the idea of a "central executive" will become redundant, and may be retired in favor of more precise descriptions of the functions needed to perform different aspect of cognition, such as goal-directed behavior and understanding language. A construct like "working memory" or "the central executive" is only useful insofar as it brings us closer to reaching the goal of understanding and explaining cognitive functions and how they are represented in and performed by the brain.

A second conclusion that can be drawn from these results concerns the most useful approach to take in trying to map cognitive functions onto the brain. Rather than using complex tests that are sensitive to multiple components of a function like working memory, it is best to create relatively simple measures of single components. Scores on complex, multicomponent tests are difficult to interpret, because individual differences in normal subjects, and impaired performance in lesioned subjects, could arise due to difficulty on any one of the test components.

II. Assessing cognitive models of working memory

The experiments described in this thesis are relevant to the debate about how best to model working memory. While SR models claim that all working memory functions rely on the same pool of resources, MC models view working memory as a set of interacting components. The fact that scores on different tests are not correlated and are sensitive to different hypothesized components (i.e., short-term store, CE) lends support to this view. In addition, the multicomponent nature of CE functions is supported by results of selective lesion studies in monkeys (Petrides, 1994; 1995) and imaging studies in humans (Petrides, 1993a, 1993b) that distinguish between brain regions used for monitoring within working memory (dorsolateral frontal cortex), conditional learning (posterior dorsal prefrontal cortex), and strategic encoding and retrieval from long-term memory (ventral prefrontal cortex).

Another distinction between SR and MC models is their explanation for individual differences in normal subjects. Individual differences have been observed in the performance of working memory tests, and in the effects of aging and neurological disease on working memory. According to SR models, individual differences on tests like Daneman and Carpenter's (1980) Reading Span measure are assumed to reflect differences in the amount of resources subjects have available to process and store information. But such models do not provide any reason why normal, healthy subjects should differ in this capacity. According to MC models, individual differences can occur for a variety of reasons, and different components of working memory may be affected by lesions to different brain reasons. In normal subjects, individual differences are most likely due to differences in the efficiency of executive processes, such as attention or coordination. Thus, individual differences may be due to differences in motivation or attention paid to the task, or differences in strategy use, rather than differences in overall capacity of working memory (Baddeley, 1996).

The experiments in Chapter 3 showed that over half the variance in sentence memory capacity was due to individual differences in working memory. N-back and Complex Span, a composite measure sensitive to short-term storage capacity, together accounted for all the unique variance contributed by the working memory measures. Word Span, a measure of short-term storage alone, also accounted for unique variance when used as a predictor in combination with N-back, suggesting that the CE and the STS make independent contributions to sentence memory capacity. The results from both chapters, showing that different working memory tests do not correlate, that more than one factor underlies the data, and that N-back and tests sensitive to storage capacity (Complex Span and Word Span, Chapter 3) make independent contributions to sentence memory capacity, provide support for MC models and contradict the predictions of SR models.

Single resource models have contributed to attempts to understand the relation between working memory and language, but have proven less valuable for forming models based on interdisciplinary data. Such models clearly originated within cognitive psychological research, without reference to results from other disciplines. Among researchers studying the neural bases of memory, it is generally accepted that memory is not a unitary system, but is made up of interacting components, and that different brain regions contribute to different components of memory (Moscovitch, 1992). While SR models have stimulated psycholinguistic investigations of the relation between sentence comprehension and working memory, such models do not provide an adequate description of working memory itself. They are unable to account for specific impairments observed following brain lesions, double dissociations between groups of subjects with different lesions, and the fact that a distributed network of brain areas are active during working memory tasks, each of which appears to contribute to different components of the task (Paulesu et al., 1993; Smith &

Jonides, 1997). Single resource models are thus inconsistent with neuropsychological and neuroimaging data, and while they may contribute to psycholinguistic theory, they do not appear to contribute to the advancement of our understanding of the relation between the brain and behavior.

III. Future Directions

The goal of generating a complete model of the cognitive and neural bases of working memory is a daunting one, but it is not entirely out of reach. The studies conducted and reviewed in this thesis lead to two broad conclusions: First, a complete model of working memory must be based on the available evidence from all relevant disciplines, including experimental studies in normal subjects, neuropsychological investigations of the effects of aging and brain lesions, and functional neuroimaging experiments. Second, tasks used to assess working memory should be pretested and validated: It is the researcher's responsibility to show the connection between the test and the function being measured. This connection may be demonstrated by presenting an analysis of the task demands, or by showing that the test is sensitive to the components of working memory that it is intended to measure. It would be extremely useful if researchers would perform such validating analyses, and then adopt some standard versions of tasks that could be used in different laboratories and with different subject groups. Without such validation and replication, it will be difficult to develop a model that will be broadly accepted outside of the laboratory in which it was generated.

References

Awh, E., Jonides, J., Smith, E. E., Schumacher, E. H., Koeppe, R. A., & Katz, S. (1996). Dissociation of storage and rehearsal in verbal working memory: Evidence from Positron Emission Tomography. Psychological Science, 7, 25-31.

Babcock, R. L., & Salthouse, T. A. (1990). Effects of increased processing demands on age differences in working memory. Psychology and Aging, 5, 421-428.

Baddeley, A. (1996). Exploring the central executive. Quarterly Journal of Experimental Psychology, 49A, 5-28.

Baddeley, A., Eldridge, M., & Lewis, V. (1981). The role of subvocalisation in reading. Quarterly Journal of Experimental Psychology, 33A, 439-454.

Baddeley, A., Logie, R., Nimmo-Smith, I., & Brereton, N. (1985). Components of fluid reading. Journal of Memory and Language, 24, 119-131.

Baddeley, A., Vallar, G., & Wilson, B. (1987). Sentence comprehension and phonological memory: Some neuropsychological evidence. In M. Coltheart (Ed.), Attention and Performance XII (pp. 509-529). Hove, UK: Lawrence Erlbaum.

Baddeley, A., & Wilson, B. (1988). Frontal amnesia and the dysexecutive syndrome. Brain and Cognition, 7, 212-230.

Baddeley, A. D. (1983). Working Memory. Philosophical Transactions of the Royal Society of London, B 302, 311-324.

Baddeley, A. D., Bressi, S., Sala, R. D., Logie, R., & Spinnler, H. (1991). The decline of working memory in Alzheimer's disease: A longitudinal study. Brain, 114, 2521-2542.

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. H. Bower (Ed.), The Psychology of Learning and Motivation (Vol. 8, pp. 47-89). New York: Academic Press.

Baddeley, A. D., & Hitch, G. J. (1994). Developments in the concept of working memory. Neuropsychology, 8, 485-493.

Baddeley, A. D., Logie, R., Bressi, S., Sala, R. D., & Spinnler, H. (1986). Dementia and working memory. Quarterly Journal of Experimental Psychology, 38A(603-618).

Basso, A., Spinnler, H., Vallar, G., & Zanobio, M. E. (1982). Left hemisphere damage and selective impairment of auditory verbal short-term memory. A case study. Neuropsychologia, 20, 263-274.

Belleville, S., Peretz, I., & Arguin, M. (1992). Contribution of articulatory rehearsal to short-term memory: Evidence from a case of selective disruption. Brain and Language, 43, 713-746.

Bever, T. G. (1970). The cognitive basis for linguistic structures. In J. R. Hayes (Ed.), Cognition and the Development of language (pp. 279-362). New York: Wiley.

Bever, T. G., Lackner, J. R., & Kirk, R. (1969). The underlying structures of sentences are the primary units of immediate speech processing. Perception & Psychophysics, 5, 225-234.

Blauberg, M. S., & Braine, M. D. S. (1974). Short-term memory limitations on decoding self-embedded sentences. Journal of Experimental Psychology, 102, 745-748.

Blumenthal, A. L. (1966). Observations with self-embedded sentences. Psychonomic Science, 6, 453-454.

Braver, T. S., Cohen, J. D., Nystrom, L. E., Jonides, J., Smith, E. E., & Noll, D. C. (1997). A parametric study of prefrontal cortex involvement in human working memory. Neuroimage, 5, 49-62.

Bryant, F. B., & Yarnold, P. R. (1995). Principal-components analysis and exploratory and confirmatory factor analysis. In L. G. Grimm & P. R. Yarnold (Eds.), Reading and Understanding Multivariate Statistics (pp. 99-136). Washington, DC: American Psychological Association.

Caplan, D. (1972). Clause boundaries and recognition latencies for words in sentences. Perception & Psychophysics, 12, 73-76.

Caplan, D., & Waters, G. S. (1990). Short-term memory and language comprehension: A critical review of the neuropsychological literature. In G. Vallar & T. Shallice (Eds.), Neuropsychological Impairments of Short-term Memory (pp. 337-389). Cambridge: Cambridge University Press.

Caplan, D., & Waters, G. S. (1995). Aphasic disorders of syntactic comprehension and working memory capacity. Cognitive Neuropsychology, 12, 637-649.

Caplan, D., & Waters, G. S. (1996). Syntactic processing in sentence comprehension under dual-task conditions in aphasic patients. Language and Cognitive Processes, 11, 525-551.

Caplan, D., & Waters, G. S. (in press). Verbal working memory and sentence comprehension. Behavioral and Brain Sciences .

Case, R., Kurland, E. M., & Goldberg, J. (1982). Operational efficiency and the growth of short-term memory span. Journal of Experimental Child Psychology, 33, 386-404.

Chang, F. R. (1980). Active memory processes in visual sentence comprhension: Clause effects and pronominal reference. Memory & Cognition, 8, 58-64.

Cowper, E. A. (1976). Constraints on sentence complexity: A model for syntactic processing. Unpublished doctoral dissertation, Brown University.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. Journal of Verbal Learning and Verbal Behavior, 19, 450-466.

Daneman, M., & Carpenter, P. A. (1983). Individual differences in integrating informatin between and within sentences. Journal of Experimental Psychology: Learning, Memory and Cognition, 9, 561-584.

Daneman, M., & Merikle, P. M. (1996). Working memory and language comprehension: A meta-analysis. Psychonomic Bulletin & Review, 3, 422-433.

D'Esposito, M., Detre, M. A., Alsop, D. C., Shin, R. K., Atlas, S., & Grossman, M. (1995). The neural basis of the central executive system of working memory. Nature, 378, 279-281.

De Roeck, A., Johnson, R., King, M., Rosner, M., Sampson, G., & Varile, N. (1982). A myth about center-embedding. Lingua, 58, 327-340.

Dobbs, A. R., & Rule, B. G. (1989). Adult age differences in working memory. Psychology and Aging, 4, 500-503.

Engle, R. W. (1996). Working memory and retrieval: An inhibition-resource approach. In J. T. E. Richardson, R. W. Engle, L. Hasher, R. H. Logie, E. R. Stoltzfus, & R. T. Zachs (Eds.), Working Memory and Human Cognition (pp. 89-119). New York: Oxford University Press.

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 & Cognition, 18, 972-992.

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. Psychological Review, 102, 211-245.

Fiez, J. A., Raife, E. A., Balota, D. A., Schwartz, J. P., Raichle, M. E., & Petersen, S. E. (1996). A positron emission tomography study of the short-term maintenance of verbal information. The Journal of Neuroscience, 16, 808-822.

Funahashi, S., Bruce, C. J., & Goldman-Rakic, P. S. (1988). Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. Journal of Neurophysiology, 61, 331-349.

Funahashi, S., Bruce, C. J., & Goldman-Rakic, P. S. (1993). Dorsolateral prefrontal lesions and oculomotor delayed-response performance: Evidence for mnemonic "scotomas." Journal of Neuroscience, 13, 1479-1497.

Gathercole, S. E., & Baddeley, A. D. (1993). Working memory and language. Hillsdale: Lawrence Erlbaum.

Gernsbacher, M. A., Hargreaves, D. J., & Beeman, M. (1989). Building and accessing clausal representations: The advantage of first mention versus the advantage of clause recency. Journal of Memory and Language, 28, 735-755.

Gershberg, F. B., & Shimamura, A. P. (1995). Impaired use of organizational strategies in free recall following frontal lobe damage. Neuropsychologia, 13, 1305-1333.

Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. Cognition. in press.

Gibson, E. (1991). A computational theory of human linguistic processing: Memory limitations and processing breakdown. Unpublished doctoral dissertation, Carnegie-Mellon University.

Gibson, E., & Thomas, J. (1997). Processing load judgements in English: Evidence for the Syntactic Prediction Locality Theory of syntactic complexity. Unpublished MIT manuscript.

Gibson, E. & Warren, T. Discourse reference and syntactic complexity. Manuscript in preparation.

Glanzer, M., Dorfman, D., & Kaplan, B. (1981). Short-term storage in the processing of text. Journal of Verbal Learning and Verbal Behavior, 20, 656-670.

Glanzer, M., Fischer, B., & Dorfman, D. (1984). Short-term storage in reading. Journal of Verbal Learning and Verbal Behavior, 20, 656-670.

Glanzer, M., & Razel, M. (1974). The size of the unit in short-term storage. Journal of Verbal Learning and Verbal Behavior, 13(114-131).

Goldman-Rakic, P. S., & Friedman, H. R. (1991). The circuitry of working memory revealed by anatomy and metabolic imaging. In H. S. Levin, H. M. Eisenberg, & A. L. Benton (Eds.), Frontal Lobe Function and Dysfunction (pp. 72-91). New York: Oxford University Press.

Harman, H. H. (1976). Modern Factor Analysis. Chicago: University of Chicago Press.

Howard, D., & Butterworth, B. (1989). Short-term memory and sentence comprehension: A reply to Vallar and Baddeley, 1987. Cognitive Neuropsychology, 6, 455-463.

Jarvella, R. J. (1971). Syntactic processing of connected speech. Journal of Verbal Learning and Verbal Behavior, 10, 409-416.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. Psychological Review, 99, 122-149.

Just, M. A., Carpenter, P. A., & Keller, T. A. (1996a). The capacity theory of comprehension: New frontiers of evidence and arguments. Psychological Review, 103, 773-780.

Just, M. A., Carpenter, P. A., Keller, T. A., Eddy, W. F., & Thulborn, K. R. (1996b). Brain activation modulated by sentence comprehension. Science, 274, 114-116.

Kac, M. B. (1981). Center-embeddings revisited. Proceedings of the third annual conference of the Cognitive Science Society. Hillsdale: Lawrence Erlbaum.

Kimball, J. (1973). Seven principles of surface structure parsing in natural language. Cognition, 2, 15-47.

King, J., & Just, M. A. (1991). Individual differences in syntactic processing: The role of working memory. Journal of Memory and Language, 30, 580-602.

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, 376-395.

Kintsch, W., & Dijk, T. A. V. (1978). Toward a model of text comprehension and production. Psychological Review, 85, 363-394.

Klapp, S. T., Marshburn, E. A., & Lester, P. T. (1983). Short-term memory does not involve the "working memory" of information processing: The demise of a common assumption. Journal of Experimental Psychology: General, 112, 240-264.

LaPointe, L. B., & Engle, R. W. (1990). Simple and complex word spans as measures of working memory capacity. Journal of Experimental Psychology: Learning, Memory and Cognition, 16, 1118-1133.

Larkin, W., & Burns, D. (1977). Sentence comprehension and memory for embedded structure. Memory & Cognition, 5, 17-22.

Lehto, J. (1996). Are executive function tests dependent on working memory capacity. The Quarterly Journal of Experimental Psychology, 49A, 29-50.

Lewis, R. (1996). A theory of grammatical but unacceptable embeddings. Journal of Psycholinguistic Research, 25, 93-116.

Light, L. L., & A., A. P. (1985). Working-memory capacity, age, and memory for discourse. Journal of Gerentology, 40, 737-747.

Logie, R. H. (1996). The seven ages of working memory. In J. T. E. Richardson, R. W. Engle, L. Hasher, R. H. Logie, E. R. Stoltzfus, & R. T. Zachs (Eds.), Working Memory and Human Cognition (pp. 31-65). New York: Oxford University Press.

Lombardi, L., & Potter, M. C. (1992). The regeneration of syntax in short-term memory. Journal of Memory and Language, 31, 713-733.

MacDonald, M. C., Just, M. A., & Carpenter, P. A. (1992). Working memory constraints on the processing of syntactic ambiguity. Cognitive Psychology, 26, 56-98.

MacDonald, M. C. & Christiansen, M. H. (1988). Individual differences without working memory: A reply to Just & Carpenter and Waters & Caplan. Unpublished manuscript.

Martin, R. C. (1987). Articulatory and phonological deficits in short-term memory and their relation to syntactic processing. Brain and Language, 32, 159-192.

Martin, R. C. (1993). Short-term memory and sentence processing: Evidence from neuropsychology. Memory & Cognition, 21, 176-183.

Martin, R. C., & Romani, C. (1994). Verbal working memory and sentence comprehension: A multiple-components view. Neuropsychology, 8, 506-523.

McCall, (1998). Fundamental Statistics for the Behavioral Sciences, 7th ed. New York: Brooks/Cole.

McCarthy, R. A., & Warrington, E. K. (1987a). The double dissociation of short-term memory for lists and sentences. Brain, 110, 1545-1563.

McCarthy, R. A., & Warrington, E. K. (1987b). Understanding: A function of short-term memory? Brain, 110, 1565-1578.

Miller, E. K. (1998). Prefrontal cortex and the neural basis of executive functions. in press.

Miller, E. K., Erickson, C. A., & Desimone, R. (1996). Neural mechanisms of visual working memory in prefrontal cortex of the macaque. Journal of Neuroscience, 16, .

Miller, E. K., Li, L., & Desimone, R. (1991). A neural mechanism for working and recognition memory in inferior temporal cortex. Science, 254, 1377-1379.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. Psychological Review, 63, 81-97.

Miller, G. A., & Isard, S. (1964). Free recall of self-embedded English sentences. Information and Control, 7, 292-303.

Milner, B., & Petrides, M. (1984). Behavioural effects of frontal-lobe lesions in man. Trends in Neurosciences, 7, 403-407.

Milner, B., Petrides, M., & Smith, M. L. (1985). Frontal lobes and the temporal organization of memory. Human Neurobiology, 4, 137-142.

Milsark, G. L. (1983). On length and structure in sentence parsing. Cognition, 13, 129-134.

Mishkin, M., Ungerleider, L. G., & Macko, K. A. (1983). Object vision and spatial vision: Two cortical pathways. Trends in Neurosciences, 6, 414-417.

Miyake, A., Carpenter, P. A., & Just, M. A. (1994a). A capacity approach to syntactic comprehension disorders: Making normal adults perform like aphasic patients. Cognitive Neuropsychology, 11, 671-717.

Miyake, A., Carpenter, P. A., & Just, M. A. (1995). Reduced resources and specific impairments in normal and aphasic sentence comprehension. Cognitive Neuropsychology, 12, 651-679.

Miyake, A. K., Just, M. A., & Carpenter, P. A. (1994b). Working memory constraints on the resolution of lexical ambiguity: Maintaining multiple interpretations in neurtral contexts. Journal of Memory and Language, 33, 175-202.

Moscovitch, M. (1992). Memory and working-with-memory: A component process model based on modules and central systems. Journal of Cognitive Neuroscience, 4, 257-267.

Nakatani, K., & Gibson, E. Memory costs of preverbal elements: Modifiers vs. arguments. Manuscript in preparation.

Olton, D. S., Becker, J. T. & Handelmann, G. E. (1979). Hipocampus, space, and memory. Behavioral and Brain Sciences, 2, 313-332.

Olton, D. S., & Feustle, W. A. (1981). Hippocampal function required for nonspatial working memory. Experimental Brain Research, 41, 380-389.

Olton, D. S., Wenk, G. L., Church, R. M., & Meck, W. H. (1988). Attention and the frontal cortex as examined by simultaneous temporal processing. Neuropsychologia, 26, 307-318.

Passingham, R. (1993). The Frontal Lobes and Voluntary Action. Oxford: Oxford University Press.

Paulesu, E., Frith, C. D., & Frackowiak, R. S. J. (1993). The neural correlates of the verbal component of working memory. Nature, 362, 342-345.

Petrides, M. (1991). Monitoring of selections of visual stimuli and the primate frontal cortex. Proceedings of the Royal Society of London, B, 246, 293-298.

Petrides, M. (1994). Frontal lobes and working memory: Evidence from investigations of the effects of cortical excisions in nonhuman primates. In F. Boller & J. Grafman (Eds.), Handbook of Neuropsychology (Vol. 9, pp. 59-82). Amsterdam: Elsevier.

Petrides, M. (1995a). Functional organization of the human frontal cortex for mnemonic processing. Annals of the New York Academy of Sciences, 769, 85-96.

Petrides, M. (1995b). Impairments on nonspatial self-ordered and externally ordered working memory tasks after lesions of the mid-dorsal part of the lateral frontal cortex in the monkey. Journal of Neuroscience, 15, 359-375.

Petrides, M. (1996). Lateral frontal cortical contribution to memory. The Neurosciences, 8, 57-63.

Petrides, M., Alivisatos, B., Evans, A. C., & Meyer, E. (1993a). Dissociation of human mid-dorsolateral from posterior dorsolateral frontal cortex in memory processing. Proceedings of the National Academy of Sciences, 90, 873-877.

Petrides, M., Alivisatos, B., Meyer, E., & Evans, A. C. (1993b). Functional activation of the human frontal cortex during the performance of verbal working memory tasks. Proceedings of the National Academey of Sciences, 90, 878-882.

Petrides, M., & Milner, B. (1982). Deficits on subject-ordered tasks after frontal- and temporal-lobe lesions in man. Neuropsychologia, 20, 249-262.

Potter, M. C. (1993). Very short-term conceptual memory. Memory & Cognition, 21, 156-161.

Potter, M. C., & Lombardi, L. (1990). Regeneration in the short-term recall of sentences. Journal of Memory and Language, 29, 633-654.

Potter, M. C., Moryadas, A., Abrams, I., & Noel, A. (1993). Word perception and misperception in context. Journal of Experimental Psychology: Learning, Memory and Cognition, 19, 3-22.

Potter, M. C., Valian, V. V., & Faulconer, B. A. (1977). Representation of a sentence and its pragmatic implications: Verbal, imagistic, or abstract? Journal of Verbal Learning and Verbal Behavior, 16, 1-12.

Rao, S. C., Rainer, G., & Miller, E. K. (1997). Integrating what and where in the primate frontal cortex. Science, 276, 821-824.

Richardson, J. T. E. (1996). Evolving issues in working memory. In J. T. E. Richardson, R. W. Engle, L. Hasher, R. H. Logie, E. R. Stoltzfus, & R. T. Zachs (Eds.), Working Memory and Human Cognition (pp. 120-154). New York: Oxford University Press.

Saffran, E. M., & Marin, O. S. M. (1975). Immediate memory for word lists and sentences in a patient with deficient auditory short-term memory. Brain and Language, 2, 420-433.

Salthouse, T. A. (1990). Working memory as a processing resource in cognitive aging. Developmental Review, 10, 101-124.

Schumacher, E. H., Lauber, E., Awh, E., Jonides, J., Smith, E. E., & Koeppe, R. A. (1996). PET evidence for an amodal verbal working memory system. Neuroimage, 3, 79-88.

Smith, E. E., & Jonides, J. (1997). Working memory: A view from neuroimaging. Cognitive Psychology, 33, 5-42.

Smith, E. E., Jonides, J., Koeppe, R. A., Awh, E., Schumacher, E. H., & Minoshima, S. (1995). Spatial versus object working memory: PET investigations. Journal of Cognitive Neuroscience, 7, 337-356.

Stabler, E. P. (1994). The finite connectivity of linguistic structures. In C. Clifton, Jr., L. Frazier, & K. Rayner, (Eds). Perspectives in sentence processing, (pp. 303-336). Hillsdale, NJ: Erlbaum.

Stromswold, K., Caplan, D., Alpert, N., & Rauch, S. (1996). Localization of syntactic comprehension by positron emission tomography. Brain and Language, 52, 452-473.

Tanenhaus, M. K., & Trueswell, J.C. (1995). Sentence comprehension. In J. L. Miller & P. D. Eimas, (Eds.), Speech, Language and Communication (pp. 217-262). San Diego, CA: Academic Press.

Tulving, E. (1972). Episodic and semantic memory. In E. Tulving & W. D. Donaldson, (Eds.), Organization of Memory, New York: Academic Press.

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? Journal of Memory and Language, 28, 127-154.

Vallar, G., Betta, A. M. D., & Silveri, M. C. (1997). The phonological short-term store-rehearsal system: Patterns of impairment and neural correlates. Neuropsychologia, 35, 795-813.

Waters, G., Caplan, D., & Hildebrandt, N. (1991). On the structure of verbal short-term memory and its functional role in sentence comprehension: Evidence from neuropsychology. Cognitive Neuropsychology, 8, 81-126.

Waters, G. S., & Caplan, D. (1996a). The capacity theory of sentence comprehension: Critique of Just and Carpenter (1992). Psychological Review, 103, 761-772.

Waters, G. S., & Caplan, D. (1996b). The measurement of verbal working memory capacity and its relation to reading comprehension. The Quarterly Journal of Experimental Psychology, 49A, 51-79.

Waters, G. S., & Caplan, D. (1996c). Processing resource capacity and the comprehension of garden path sentences. Memory & Cognition, 24, 342-355.

Waters, G. S., & Caplan, D. (1997). Working memory and on-line sentence comprehension in patients with Alzheimer's disease. Journal of Psycholinguistic Research, 26, 377-400.

Waters, G. S., Caplan, D., & Hildebrandt, N. (1987). Working memory and written sentence comprehension. In M. Coltheart (Ed.), Attention and Performance XII (pp. 531-555). Hove, UK: Lawrence Erlbaum.

Waters, G. S., Caplan, D., & Rochon, E. (1995). Processing capacity and sentence comprehension in patients with Alzheimer's disease. Cognitive Neuropsychology, 12, 1-30.

Wilson, F. A. W., O Scalaidhe, S. P., & Goldman-Rakic, P. S. (1993). Dissociation of object and spatial processing domains in primate prefrontal cortex. Science, 260, 1955-1958.

Wingfield, A., Stine, E. A. L., Lahar, C. J., & Aberdeen, J. S. (1988). Does the capacity of working memory change with age? Experimental Aging Research, 14, 103-107.

Epilogue

Trials never end, of course. Unhappiness and misfortune are bound to occur as long as people live, but there is a feeling now, that was not here before, and is not just on the surface of things, but penetrates all the way through: We've won it. It's going to get better now. You can sort of tell these things.

- Robert Pirsig,
Zen and the Art of Motorcycle Maintenance