The Inner Workings of Working Memory:
The effects of aging and language impairment on tasks examining verbal working memory

by

Rebecca A. Hayes

Submitted to the Faculty of

School of Health and Rehabilitation Sciences in partial fulfillment

of the requirements for the degree of

Bachelor of Philosophy

University of Pittsburgh

2011

UNIVERSITY OF PITTSBURGH

SCHOOL OF HEALTH AND REHABILITATION SCIENCES

This thesis was presented

by

Rebecca A. Hayes

It was defended on

April 11, 2011

and approved by

William Hula, Ph.D., Adjunct Assistant Professor, Communication Science and Disorders, VA Pittsburgh Healthcare System

Tessa Warren, Ph.D., Associate Professor, Psychology, Linguistics, & Communication Science and Disorders

External Examiner: Heather Harris Wright, Associate Professor, Speech and Hearing Science, Arizona State University

Thesis Director: Michael Walsh Dickey, Assistant Professor, Communication Science and Disorders, VA Pittsburgh Healthcare System

The Inner Workings of Working Memory:
The effects of aging and language impairment on tasks examining verbal working memory

Rebecca A. Hayes

University of Pittsburgh, 2011

Wright et al. (2007) tested Persons with Aphasia (PWA) using three N-Back tasks featuring different types of linguistic information – phonological, semantic, and syntactic -- to determine whether Verbal Working Memory (VWM) is a single, united resource. The current study tested three PWA with the same tasks, as well as an additional vision-focused task, to expand on this previous research; two groups of cognitively normal individuals were tested using the same protocol to provide a baseline for comparison. Results from the unimpaired groups indicated no effects of aging, and significant differences in performance across all types of information except phonological and visual cues. Results from PWA were inconclusive. The N-Back task, however, was found to cause misleading patterns in accuracy scores for some tests; sensitivity scores are suggested as a better measure of performance on this testing paradigm.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

This thesis was made possible by the help of many. The following people and institutions have my sincere gratitude:

The Brackenridge Summer Research Fellowship, which provided funding for this project and a community of undergraduate researchers whose insights were invaluable during the planning and execution of this project.

The University Honors College, which offers many resources and opportunities for undergraduate researchers.

The Western Pennsylvania Patient Registry, for helping us recruit persons with aphasia.

The Osher Lifelong Learning Community, for helping us recruit unimpaired adults.

My committee, for their time and suggestions, and for making the defense of this thesis possible:

- Dr. Michael Walsh Dickey, my thesis advisor. Words cannot describe the patience and enthusiasm that Dr. Dickey has shown throughout this project. This paper would not exist without him.
- Dr. Heather Harris Wright, my external examiner, for providing the inspiration for this project and sending her original tests, and for travelling from Arizona to be part of the committee.

- Dr. Tessa Warren and Dr. William Hula, who have offered their support and resources throughout the duration of this study.

Dominique Hensley, for her incredible help and support while testing persons with aphasia.

Finally, my family and friends, for being so supportive and encouraging as I navigated the research process for the first time.

1.0 INTRODUCTION

Aphasia is a condition caused by damage to the language centers of the brain, most typically by a cerebrovascular accident [CVA, also known as a 'stroke'] (NINDS, 2009). Aphasia can affect language in many different ways, one of which is impaired comprehension of written or spoken language. The condition varies in severity and symptoms, often based on the area of the brain in which the damage occurs. The exact point in processing at which comprehension breaks down for people with aphasia is still unclear (Dickey & Thompson, 2004); theories range from a deficit in structural 'rules' to slowed lexical activation to reduced working memory [WM] capacity. Even within the theory of WM reduction there is debate – are all WM resources reduced or solely those used for language processing? Research during recent decades on WM and aphasia sought to answer these questions about the nature of the disorder, as well as to learn more about the structure of WM itself.

Working memory is acknowledged by many experts as a crucial factor in language comprehension (Caplan & Waters, 1999; Just & Carpenter, 1992), but opinions differ as to the exact nature of this cognitive resource. An older theory developed by Baddeley divided WM into a language-centric 'phonological loop' and a 'visuo-spatial scratchpad', which are connected by a central executive process (Caplan & Waters, 1996). Currently, the psycholinguistic community is divided as to whether VWM is even further segmented into separate resources for different types of linguistic processing, or whether WM is one united cognitive resource from which all processes draw. Furthermore, while some theories on the nature of aphasia have included reduced WM resources, little research exists comparing the WM of aphasic persons to their typically-functioning contemporaries.

This study is an attempt to lend information to both of these issues – the nature of WM itself, and its status in the aphasic population. WM measures are usually preliminary tests in research, used to divide participants into high- and low-capacity groups or to establish correlations between WM span and performance on some other task. The goal of this study, however, is to use a WM measure to gain normative data from each of the three participant populations, which can then be compared both within and between groups.

1.1 APHASIA

Recent estimates indicate that about one million Americans currently have aphasia (NIDCD 2008, NINDS 2010). According to the National Institute on Deafness and Communication Disorders [NIDCD], "[a]phasia is a disorder that results from damage to portions of the brain that are responsible for language," (NIDCD 2008). Several of these portions have been identified since Broca's breakthrough research in 1865 (Geschwind, 1972). Lesions in different areas of the brain have traditionally been associated with different types of aphasia, characterized by different language difficulties (Geschwind, 1972). The two most well-known types of aphasia include Broca's, a type of "nonfluent" aphasia, which is most known for patients' slow and labored speech paired with reduced ability to use grammatical words and phrases, and Wernicke's, a type of "fluent" aphasia, wherein patients' speech is rapid but holds little semantic content, and patients have severe reductions in language comprehension (Geschwind, 1972).

Reductions in syntactic comprehension ability, however, have been demonstrated regardless of lesion site (Caramazza & Zurif, 1976).

Explanations for these comprehension difficulties vary, but can generally be separated into two basic theories: either PWA lack the cognitive resources needed to parse sentences effectively, or they lack the requisite strategies (Caplan et al., 2007). The theory this study focuses on, that of reduced Working Memory resources, falls firmly into the former category.

1.2 WORKING MEMORY

1.2.1 Modular vs. Constraint-Based Theories: Two Landmark Papers

One of the most debated issues in the study of language comprehension is the nature of working memory [WM]. In the 1980s, Baddeley and colleagues conducted a series of studies that suggested all cognitive WM was modular, divided into three parts: the phonological loop, the visual-spatial scratchpad, and the central executive (Caplan & Waters, 1999). Generally, researchers assume that the resources used in processing language are separate from those used in other types of processing, but there is unrest in the field over what exactly verbal WM comprises. Is there one communal pool of resources from which language processing draws? Or is WM broken down into smaller parts, allocated to specific elements of storage and processing? Two teams of researchers have written landmark articles on this issue – Caplan and Waters, and Just and Carpenter.

1.2.1.1 Just & Carpenter: Capacity-Constrained Theory

Just and Carpenter (1992) argue for a capacity-constrained theory of language comprehension, where an individual's WM span determines his or her comprehension abilities. While most research in the preceding century had focused on storage span alone, Just and Carpenter argued that WM controls processing as well as storage. This hypothesis was based on research done by Baddeley and Hitch that compared storage tasks to retrieval tasks; Just and Carpenter, however, argue for one resource pool from which both draw, as opposed to Baddeley and Hitch's separation of the two into the phonological loop and central executive.

Just and Carpenter investigated the idea that activation levels of linguistic information are the driving force behind cognitive processes of storage and computation. The more WM capacity an individual has, the more elements that can be activated while simultaneously completing language processing. To test this idea, the investigators divided subjects into high- and low- capacity groups using a standard word-retention task developed by Daneman and Carpenter. This division allowed the data from several experiments to be analyzed for individual differences in sentence processing as related to WM capacity.

Just & Carpenter describe several studies in which reading time measures were taken while participants read sentences with different, specific syntactic structure. These structures included “syntactically ambiguous” sentences – which were first investigated to see whether semantic content would affect parsing and then examined to see whether parallel interpretations of each possible parse were being constructed. Just and Carpenter found that higher-span subjects exhibited single-resource-type behavior, taking semantic content into account during parsing and constructing parallel parses of ambiguous sentences, whereas lower-span subjects did not. Just and Carpenter interpret these results to mean that WM span and syntactic processing are connected.

While this paper provides convincing evidence for a non-modular, single-resource theory of WM, some concerns arise. All of these experiments used reading-intensive paradigms, which are generally considered less natural than aural comprehension tests. Additionally, Caplan and Waters (1999) point out that the Daneman and Carpenter test used to separate subjects into high- and low-span groups can be unreliable. The test does not measure reaction times, only accuracy, so the storage spans acquired by this test may be caused by trade-offs between processing and storage attention allocation.

1.2.1.2 Caplan & Waters: Modularity Beyond Baddeley

By contrast, Caplan and Waters (1999) argue for a more modular-type system than Baddeley, investigating the possibilities of a "separate-sentence-interpretive-resource," (p. 79) [SSIR]. While Just and Carpenter's single-resource theory predicts that low WM would reduce processing resources, Caplan and Waters' theory states that general storage WM tasks would not predict processing ability. The authors assert that correlations between storage WM measures and processing performance may be caused by low-span subjects struggling with other demands of WM span tasks.

One of the most crucial predictions of SSIR theory is that external memory loads will not interfere with sentence-internal processes; this is in direct opposition to the single-resource [SR] theory that external memory loads will leach resources from the internal computations. Caplan and Waters describe several studies which implemented extrinsic loading while participants performed auditory sentence-processing tasks. Caplan and Waters found none of the behaviors that Just & Carpenter would have predicted - sentence parsing showed no supra-additive effects from WM loading tasks.

Caplan and Waters describe a large amount of evidence for their SSIR theory, but they acknowledge that some of it is weak. Much of their evidence against single-resource model is based on the *absence* of a result. Unless the sample sizes of the studies are very large, null results are considered very unreliable. Additionally, a lot of the studies presented in this paper were replications that did not reproduce the original results. This situation calls into question the reliability of the measures used.

1.2.2 Continuing Dissent

While the two articles described above are considered staples in any discussion of VWM, many other authors have provided evidence on both sides of the modular/single-resource argument. Nation et al. (1999) asserted that, in children, reduced verbal WM resources are at least partially responsible for poor reading comprehension. Gibson (2000) proposed a system of determining how much a sentence will tax WM, and thus predict performance – the more WM required to parse a sentence, the more difficult it will be to comprehend. These arguments indicate a single-resource theory for all language comprehension, in opposition to Caplan and Waters's SSIR theory. Nation, however, also argued that verbal WM capacity is not connected to other WM; in the study, children with poor verbal WM performed as well as children with normal verbal WM on a visuo-spatial WM task, suggesting some level of modularity to the WM system.

1.3 N-BACK TASK

Currently, one popular test of WM is the “N-back task,” (Wright et al. 2007). Participants listen to a string of items – words or sentences – and indicate when an item matches one they heard a specified number of items earlier. Three important variables can be controlled using the N-back paradigm: the type of information being presented, the “difficulty” or length of time that information must be maintained, and the level on which the information must match – e.g. “identity” tasks, where the participant looks for a target to be exactly the same as its predecessor, and “depth” tasks, where the target should simply share some specified feature with the previous item. For example, in a phonological 2-back depth task, a subject might hear a string that included ‘snake… cat… deer… bat.’ Since ‘cat’ and ‘bat’ rhyme, and ‘cat’ appeared two items before ‘bat,’ the participant would indicate that he or she had heard a match. A participant’s span is determined by how far apart the items can be before the individual no longer recognizes their relationship.

1.3.1 Validity: Miller et al. (2009)

Miller et al. (2009) investigated the validity of N-back tasks by testing subjects with and without Parkinson's Disease using 0-, 1-, 2-, and 3-back visual letter recognition tasks as well as several neuropsychological evaluations. These evaluations included:

1. the Mini-Mental State Examination - a brief questionnaire that covers orientation, memory, attention, verbal ability, and spatial awareness typically used to screen for dementia (Folstein et al., 1975),

2. the Digit Span subtest of the WAIS-III - a test that asks subjects to store and retrieve a string of digits (Miller et al., 2009),
3. the Stroop Interference Task - three tests, where subjects first read color names, then identify the color of ink some non-words are printed in, then identify the color ink in which a different color name was printed,
4. the Trail Making Test part A - a task where subjects connected numbers scattered across a paper in increasing numerical order as fast as possible, and
5. The Geriatric Depression scale - a questionnaire aimed specifically at older adults to identify "nonsomatic" symptoms of depression.

These five well-established neuropsychological tests were then compared with the n-back tasks, in order to determine whether any correlation could be found between this purported measure of working memory and any cognitive process.

Reaction times and accuracies for all n-back tasks were normally distributed. No correlations between the RTs or accuracies were found for most of the neuropsychological tests, but a significant correlation between the 2-back accuracy scores and the TMT-A results was found. These results suggest that some element used in completing the 2-back task is involved in the processing used during the TMT-A; the absence of a correlation between any of the other n-back tasks and the TMT-A suggests that this element is unique to the 2-back condition.

1.4 FURTHER MODULARITY IN VWM

1.4.1 Martin et al. (2003)

Martin et al. (2003) used fMRI scanning to examine regions of activation in the brain during two different types of short-term memory tasks: phonological and semantic. The test protocol involved first showing the participant either one or four written words. After a specified number of seconds, a probe word was displayed. The 11 neurally unimpaired participants were asked to judge whether the word rhymed with [in the phonological test] or was a synonym to [in the semantic test] the/a previously presented word. Participants were encouraged to rehearse the original word[s] sub-vocally during the phonological condition, but were discouraged from doing so on the semantic condition. Participants were scanned using an fMRI machine to determine whether these different types and sizes of memory loads would activate different parts of the brain. Accuracy scores were also obtained.

Similar accuracy scores suggested that the two task types were similar in difficulty. An analysis of variance performed on the results of the fMRI scans showed that there was a significant main effect of load - the longer lists did activate the brain in general more than the one-word stimuli. No activation level effects were seen for task type, and no interaction between task type and load size were found. Analyzing separate regions of interest in the brain, however, revealed an effect of task: phonological stimuli activated a portion of the parietal region more strongly than semantic stimuli. No region of the brain was found that was more activated by semantic stimuli than phonological.

These results suggest that the phonological task is different from the semantic task in some way, as different areas of the brain are more stimulated by the phonological task than the semantic one. The researchers acknowledge that subjects could have performed the semantic task using the same strategies as the phonological task - that is, subvocal rehearsal - despite having been instructed not to. If this problem occurred, it could have masked further evidence of differences as the semantic task was treated more like a phonological one, preventing possible semantic 'centers' from activating. The researchers conclude that more evidence, both neurological and behavioral, is needed to develop a task that eliminates this possibility.

1.4.2 Friedmann & Gvion (2003)

Friedmann & Gvion (2003) examined even further divisions within VWM. The authors investigated the relationship between WM deficits and comprehension of sentences that require reactivation of different types of information. The study included three subject groups: cognitively unimpaired controls, conduction aphasics, and agrammatic aphasics. These subjects were assessed using several different tasks to determine their WM span for several different information types, including digit span, word span, and non-word span. Participants then completed a sentence-comprehension task, where two types of reactivation were tested: phonological, and semantic/syntactic.

Semantic/syntactic reactivation was tested using a universally-acknowledged stumbling block for agrammatic aphasics: relative clauses. These types of sentences require a reactivation of the subject at some point later in the sentence. For example, the sentence "This is the woman that the man kissed" requires a reactivation of "woman" as an object after the verb "kissed," to complete the transitive verb phrase. This type of relative clause, called object-relative, is generally harder for agrammatic aphasics to parse than a subject relative clause, e.g. "This is the woman that kissed the man." In the subject-relative structure, the reactivation retains the antecedent's semantic role as subject of the verb, instead of forcing the comprehender to change the antecedent's semantic role to the object of a verb while reactivating it.

Phonological reactivation was tested using a somewhat less familiar setup: a sentence which, in the beginning, strongly biases interpretation of a temporary lexical ambiguity toward one meaning, but resolves to be the other meaning at the end. Temporary lexical ambiguities are caused by homophonic pairs - for example, in the sentence "The pen that the man received from his wife was packed with sheep," the word 'pen' could initially mean either a writing implement or a fenced-in area. Near the beginning of the sentence, the reader is biased to interpret 'pen' as having the former meaning, as that sort of pen is a common gift to a retiree. The end of the sentence, however, makes it clear that 'pen' in this case has the second meaning.

The researchers manipulated both the type of information being reactivated, and the distance between the antecedent and its reactivation in the sentence. Agrammatic aphasics, who were found to have limited recall-based WM (as opposed to recognition-based WM), performed worse on object-relative than subject-relative clauses, regardless of distance between antecedent and reactivation. These participants also performed poorly on the phonological reactivation sentences regardless of distance. Conduction aphasics, who demonstrated limited WM regardless of information type or modality (recall vs. recognition), demonstrated normal comprehension of all sentences with relative clauses, regardless of distance between antecedent and reactivation. Intriguingly, these participants' comprehension of the phonological reactivation sentences did decline as the distance between antecedent and reactivation increased.

The pattern of the conduction aphasics' results on the phonological reactivation task suggests that phonological WM is different from the kind of WM needed to parse complex sentences, because if these tasks used the same resources the participant group should have shown distance effects either on both or on neither of the tasks. Furthermore, syntactic WM seems to be intact for persons with conduction aphasia despite their poor performance on span tasks, but is clearly impaired for persons with Broca's aphasia in spite of their having performed better on certain span tasks. This differential reduction in performance is a further indicator that there are separate resources at work.

1.4.3 Wright et al. (2007)

Wright et al. (2007) asserted that people with aphasia [PWA] have Working Memory deficits, and acknowledged the two different schools of thought about the nature of working memory: either it is a single 'pool' of memory resources, as put forth by Just & Carpenter (1992), or it is separate abilities for distinct processes, as argued by Caplan & Waters (1999). The latter hypothesis is mentioned as a possible explanation for the results of a preliminary study performed by Downey and colleagues, which tested semantic WM and found that it bore no relationship to sentence comprehension. Wright and colleagues sought to investigate this modular-type school of thought by devising tasks that tested different areas of language processing. Furthermore, if these tasks confirmed the separate-abilities hypothesis, Wright et al. questioned which areas of language processing would see the poorest performance in PWA. Identifying where the breakdown in processing occurs in PWA would be exceedingly useful in understanding and developing treatments for aphasia.

Wright et al. selected nine participants, of which five had anomic aphasia, two had Broca’s aphasia, and one each had conduction and transcortical motor aphasia. An earlier experiment demonstrated that n-back tests are appropriate for use in studying these aphasic populations. The authors then developed n-back tasks to test three different types of processing:

phonological, semantic, and syntactic. In the PhonoBack test, participants listened for words that rhymed; in the SemBack test, they listened for words in the same category - fruit, tools, furniture, animals, or clothing; in the SynBack test, they listened for sentences of the same structure, either active or passive. These three types were further divided by task difficulty - either 1-back (so that the target unit immediately followed its matching antecedent), or 2-back (so that the target unit and antecedent were separated by one unit). Finally, the PhonoBack and SemBack tasks were divided into two different processing levels - identity, where the target and antecedent were the same word, and depth, where the target and antecedent simply rhymed or were of the same category, respectively. The SynBack depth level was deemed too difficult for the patient population and was dropped, leaving only the identity level in which the sentences were identical.

In addition to the n-back tests, participants completed a SOAP test, which measures comprehension of sentences with varying levels of complexity. Wright and colleagues compared this measure of sentence comprehension to scores on the three n-back tests, in order to determine whether any one type of verbal WM most effectively predicted comprehension scores. The authors predicted that the SynBack test would show the strongest correlation to the SOAP.

The results were reported in terms of "hit rates," or number of times that participants correctly identified one of the 20 target items. Analysis of Variance (ANOVA) revealed significant main effects of test type and n-back level. Results of planned comparisons did not, however, show any significant difference in scores on the three n-back tests; while most participants performed better on the SemBack task than the others, the difference was not statistically significant when family-wise error was controlled for. Participants' performance did, however, support the hypothesized connection between syntactic WM and sentence comprehension. Participants who did poorly on the SynBack task also scored poorly on the non-canonical items of the SOAP test, suggesting that syntactic WM is related to comprehension, at least in sentences traditionally considered to be more ‘complex.’

This experiment is notable for its distinction between different types of language information and the possible division of WM to correlate to these types, presenting a model of WM that is even more modular than Caplan & Waters’ storage/processing division. The results provide some evidence for this idea – a significant main effect of task type was found. Planned comparisons, however, show no significant differences in performance between tasks. The relatively small and varied sample may have affected the significance of the results. A larger sample size may yield stronger results. Additionally, there is no cognitively-normal baseline with which to compare the performance of the PWA in this study. With no baseline, it is possible that the relative difficulty of the tasks differs, and any differences between tasks PWA demonstrate could be caused by differences between tasks. These two additions to the protocol could confirm or refute the presence of VWM deficits in PWA.

2.0 CURRENT STUDY: GOALS AND QUESTIONS

The current study aims to expand upon Wright et al.'s findings, by comparing the performance of PWA to that of cognitively unimpaired controls. The specific research questions are as follows:

1. What is the nature of Working Memory?
   - Is there one resource for all verbal information, or separate resources for different types of verbal input?
2. What role, if any, does Verbal Working Memory play in the difficulties PWA experience?
   - If WM is affected, is it universal or limited to VWM?
   - Do PWA show differential WM performance between types of verbal tasks?

To answer these questions, participants completed a subset of Wright et al.'s language-related n-back tasks, as well as an original ShapeBack task developed for this study.

In comparing the performance of PWA to cognitively unimpaired individuals, several patterns are possible. If aphasic subjects perform significantly worse on any one of the language tasks as compared to typical adults, a theory that indicates separate WM resources for different types of language processing would be supported. Alternatively, if aphasic subjects perform worse across all three language tasks as compared to their normally-functioning counterparts, a more general language resource model would be indicated. While both the aforementioned outcomes would support a WM-related cause for aphasia, no significant difference in performance between aphasic participants and typical adults would suggest that other processing deficits are at fault.

If aphasic subjects perform similar to typically-functioning adults on the shape task, while lagging on the language-related tasks, a Baddeley-like model with separate linguistic and spatial resources is suggested. In such a model, only the linguistic WM of people with aphasia would be affected. If, instead, aphasic participants show lower performance on the shape task in addition to the language tasks, a unified reduction in working memory for aphasic patients would be implied.

2.1 HYPOTHESES

Given the body of research indicating modularity within VWM (Caplan & Waters, 1999; Friedmann & Gvion, 2003; Martin et al., 2003), and a connection between WM capacity and sentence comprehension (Caplan & Waters, 1999; Just & Carpenter, 1992; Wright et al., 2007), it is predicted that PWA will show reductions in performance on the different types of tasks; specifically, PWA are expected to show an increased reduction in performance on the SynBack task as compared to the other two language-based tasks. By contrast, because aphasia is language-specific, PWA are expected to show no difference in performance from the unimpaired controls on the ShapeBack task. Unimpaired controls are expected to perform at ceiling rates of accuracy for all tasks; no significant differences in performance are expected between the two unimpaired participant groups. Thus, any difference between performance of PWA and unimpaired controls is expected to be due to reductions from aphasia, not task difficulty or age.

This prediction should be supported by poor n-back performance corresponding to poor performance on language screening tasks.

3.0 METHOD

3.1 PARTICIPANTS

Participants were divided into two main groups: PWA and cognitively unimpaired. Members of both groups were required to have normal or corrected-to normal vision and hearing, were between the ages of 18 and 90, and had no history of neurological, neuropsychological, or neuropsychiatric conditions that could cause language problems. Cognitively unimpaired participants had no history of language disorders; participants with aphasia had no history of language disorders prior to their current condition. Additionally, participants with aphasia were required to have acquired their current condition from a left-hemisphere CVA.

3.1.1 Recruitment

Cognitively unimpaired participants were recruited in two ways: through the Research Participant Registry (PRO08010419), and through the use of advertisements (flyers and posters, in both physical and electronic form). The Research Participant Registry is a voluntary database of individuals who are willing to consider participating in research studies. This database was used to recruit adults outside the usual age range of the University of Pittsburgh student body. Advertisements were distributed throughout the general community by the PI, and to the student body at the University of Pittsburgh by the PI and the Osher Lifelong Learning Institute.

Participants with aphasia were recruited using the Western Pennsylvania Patient Registry (PRO07080061). The Western Pennsylvania Patient Registry (WPPR) is a registry of language-impaired adults who have consented to be contacted regarding possible participation in research studies which may be of interest to them.

All participants underwent further screening; of the 42 unimpaired individuals and 6 PWA screened, 30 and 3 were admitted, respectively.

3.1.2 Demographics

The cognitively unimpaired participants were divided into two subgroups: those over 50 years of age, and those under 30. The mean age of participants in the older participant group at the time of testing was 62 years 6 months, ranging from 50 years 3 months to 70 years 9 months. Members of this group were residents of the greater Pittsburgh area, or residents of relatively small communities near Youngstown, Ohio. 13 members of the Older Control group were right-handed; 2 were left-handed.

The mean age of the younger participant group was 20 years 9 months, with a range of 18 years 10 months to 22 years 5 months. Participants under 30 were mainly recruited from the University of Pittsburgh campus, with only one resident of an Ohio community in this group. All members of this group were right-handed.

Participants with aphasia were residents of the greater Pittsburgh area. The mean age of these participants was 61 years 3 months, with a range from 47 years 8 months to 68 years 0 months. All members of this group were right-handed before their strokes. Detailed profiles on each participant with aphasia can be found in Table 1.

Table 1: PWA Profiles

Subject	Age at Testing	Premorbid Handedness	Aphasia Type & Severity	Raven's	PALPA 2* Phono. processing		PALPA 47* Semantic processing	Sent type	PALPA 55* Syntactic processing	Nback perf**
PWA 1	68 Years, 0 Months	R	Type: Anomic	31	Same	36	40	reversible(20)	15	Phono: NSD
			AQ:77.0		Diff	35		non-reversible(16)	15	Sem: NSD
								gap as subj(8)	7	Syn: NSD
								gap not as subj(8)	5	Shape: NSD
								converse relations(8)	5	
PWA 3	47 Years, 8 Months	R	Type: Broca's	32	Same	36	37	reversible(20)	17	Phono: SD
			AQ: 55.2		Diff	31		non-reversible(16)	13	Sem: MSD
								gap as subj(8)	7	Syn: NSD
								gap not as subj(8)	4	Shape: MSD
								converse relations(8)	5	
PWA 6	68 Years, 7 Months	R	Type: Broca's	34	Same	32	38	reversible(20)	14	Phono: SD
			AQ: 48.2		Diff	31		non-reversible(16)	11	Sem: SD
								gap as subj(8)	6	Syn: SD
								gap not as subj(8)	6	Shape: SD
								converse relations(8)	7	

* Scores that are not within normal limits are highlighted

** As determined by single-subject analysis: SD = significant difference from controls, NSD = no significant difference, MSD = marginally significant difference

3.2 MATERIALS

3.2.1 Screening Tests

All participants completed a researcher-developed basic medical history questionnaire, and underwent a pure-tone bilateral hearing screening at 40dB using a standard audiometer with over-the-ear headphones. All participants enrolled in the study passed the hearing screening. Furthermore, all participants enrolled in the study reported English as their primary language, and reported no prior history of hearing disorders or neurological, neuropsychological, or neuropsychiatric conditions that could cause language problems (except for CVA for the participants with aphasia).All participants also completed Raven's Coloured Progressive Matrices (RCPM), with responses recorded on a researcher-developed form. The RCPM is used to test nonverbal reasoning ability (Raven, 1965). All participants enrolled in the study made 11 or fewer errors, which is within the range reported for a sample of cognitively unimpaired adults by Kertesz (1979).

3.2.1.1 PWA Screening: WAB, ABCD, & PALPA

Participants with aphasia completed additional testing to analyze their language and cognitive abilities:

1. Western Aphasia Battery-Revised (Kertesz, 2007): The WAB is a standardized test of language function in PWA. This test analyzes the type and severity of the examinee's

aphasia, through a series of subtasks that involve both expressive and receptive language. While much of the WAB can be scored on-line using the provided forms, the expressive tasks were recorded for further analysis after testing was concluded using a Dell Inspiron laptop.

2. Arizona Battery for Communicative Disorders of Dementia (Bayles & Tomeoda, 1993): Participants completed the story retelling subtasks of the ABCD. These tasks require the examinee to listen to a story and repeat it – immediately after the story was originally told, and again after a specified interval of time. This test was used to measure language-related memory decline among PWA. This test was also scored using the provided forms and recorded for further analysis.
3. Psycholinguistic Assessment of Language Processing in Aphasia: Participants also completed three subtests of the PALPA – 2, 47, and 55. These subtests were selected to relate to the three language subtasks of the experiment. Subtest 2 requires the examinee to identify whether minimal pairs of words are the same or different, a skill called upon by the PhonoBack task of the experiment. In subtest 47, participants matched pictures to words spoken by the examiner. This activity tests the semantic comprehension of the examinee, which is pertinent to the SemBack task. Finally, in subtest 55, examinees matched various types of sentences to pictures. This task tests syntactic comprehension, which is relevant to the SynBack task in the experiment.

All three of the PALPA subtests were scored using the provided forms.

3.2.2 Experimental Task

The language subtasks of the N-Back test used in this experiment are identical to third and fourth blocks of the 'identity-level' tasks used by Wright et al. (2007). A shape subtask was developed using five abstract non-nameable shapes created using the method outlined by Hautzel et al. (2009) and following the same overall outline of the Wright et al. tasks (Appendix A). All four subtasks were preceded by a practice block, containing five items and one target. Samples from these tasks are illustrated in Table 2.

Table 2: Samples from Experimental Tasks

Test Type	1-Back: Sample	2-Back: Sample
PhonoBack	pat, hat, mat, **mat**, rat	rat, pat, bat, rat, **bat**
SemBack	grape, lemon, lime, orange, **orange**	grape, apple, orange, lemon, **orange**
SynBack	The singer was blamed by the teacher The teacher blamed the singer The singer was called by the actor **The singer was called by the actor** The golfer called the jogger	The teacher blamed the lawyer The baker was chased by the golfer The singer was hugged by the jogger **The baker was chased by the golfer** The golfer was blamed by the banker
ShapeBack		

*** Target words and sentences are indicated by underscoring and bolding; target shapes are indicated by outlines**

The test was programmed and run using Psychology Software Tools's E-Prime software on a Dell 4500 series desktop computer. Verbal stimuli were presented via padded over-the-ear headphones, and visual stimuli via a flat-screen monitor. Participants responded using the space bar on a standard US keyboard, which was marked with blue tape for easy identification. Participants' accuracy and response time for each item were recorded by the E-Prime software, and later imported to Microsoft Excel and SPSS spreadsheets for data analysis.

3.3 PROCEDURE

3.3.1 Screening

As part of screening procedures, participants completed a questionnaire asking about their personal medical history, handedness, language status (i.e. whether they are native speakers of English), and vision status. All participants also underwent a short hearing screening of pure tones at 500, 1000, 2000, and 4000 Hz at 30 dB. Finally, all participants completed Raven's Coloured Progressive Matrices. Participants with aphasia also completed the additional screening tasks outlined above.

3.3.2 Experimental Task

Once the screening procedures were complete, the participants were given verbal instructions for the N-Back test and completed a short practice session as outlined in Appendix B. Once the instructions were understood and the practice items mastered, participants began the experimental tests.

4.0 RESULTS

4.1 ANALYSIS AND DESIGN

This was an experimental study using a cross-sectional design. Three participant groups – language-impaired, older cognitively normal, and younger cognitively normal – experienced four different test conditions - semantic, phonetic, syntactic, and visual – and two difficulty levels – 1-Back and 2-Back. Within-subject variables independent variables thus included the four task types and two difficulty levels of each task, and the between-subjects independent variable was membership in one of the participant groups. The n-back program recorded two dependent variables: subjects' accuracy and reaction times for each trial. The accuracy data was used to calculate the additional dependent variables of Criterion and D Prime sensitivity scores. Younger cognitively normal subjects acted as the control, while the older cognitively normal subjects will allow comparison for possible aging effects during data analysis of participants with aphasia.

4.2 YOUNGER AND OLDER CONTROLS

4.2.1 Accuracy Data

An Analysis of Variance [ANOVA] was performed on the data, revealing no significant main effect of group ($F[1,28]=0.564$, $p>.4$) and no significant interactions. There was, however, a significant main effect of difficulty level ($F[1,28]=39.70$, $p<.001$), with both groups showing generally poorer performance on the 2-back difficulty level. A significant main effect of task type ($F[3,26]=78.23$, $p<.001$) was also found; both groups performed the poorest on the SynBack task.

The cognitively unimpaired participants, regardless of age group, were expected to show a decline in accuracy from the 1- to 2-back condition, regardless of task type, because of the greater demands of the second task. Paired-samples t-tests show that for most of the tasks this prediction is true. Both the younger and older unimpaired groups, however, showed a significant *increase* in accuracy from 1- to 2-back in the SynBack task (Younger Adults: $t(14) = -3.838$, $p<.005$; Older Adults: $t(14) = -3.136$, $p<.01$). These results are shown in Table 3; instances where the difference between 1- and 2-back was significant ($p < .05$) are highlighted. As Table 3 shows, when the data from the Younger and Older age groups are combined, accuracy scores show a significant decrease from the 1- to 2-back condition in every task type except SynBack, which shows a significant increase.

This unexpected increase in accuracy is misleading, as shown by the d' sensitivity measures included in Table 3. When participants' overall sensitivity to the signal is analyzed, instead of raw accuracy, performance does not increase.

Table 3: Unimpaired Adults' Accuracy and d' Scores

Younger Unimpaired Adults								
	ShapeBack		PhonoBack		SemBack		SynBack	
	ACC	d'	ACC	d'	ACC	d'	ACC	d'
1-back	.94948713	4.271333	.92948713	3.701667	.98158973	5.192667	.74651280	2.108667
2-back	.92473687	3.358000	.91708773	3.127800	.95950880	4.775133	.77863153	1.699333
Older Unimpaired Adults								
	ShapeBack		PhonoBack		SemBack		SynBack	
	ACC	d'	ACC	d'	ACC	d'	ACC	d'
1-back	.97117947	4.437133	.91035887	3.148867	.98687180	5.366267	.74461533	1.868467
2-back	.90743867	3.218533	.87631573	2.803467	.95578940	4.263600	.77449127	1.490133
All Cognitively Unimpaired								
	ShapeBack		PhonoBack		SemBack		SynBack	
	ACC	d'	ACC	d'	ACC	d'	ACC	d'
1-back	0.9603333	4.3542333	0.919923	3.4252667	0.98423077	5.279466	0.74556407	1.988566
2-back	0.91608777	3.2882666	0.89670173	2.9656333	0.9576491	4.519366	0.7765614	1.594733

A paired-samples t-test revealed that, for younger participants, there was no significant difference between d' measures for PhonoBack and ShapeBack at the 1-back level (t = -1.356, p >.19). For both participant groups, PhonoBack and ShapeBack at the 2-back level showed no significant difference (Younger: t = -.575, p>.5; Older: t = -1.162, p>.27). All other comparisons showed significant differences in sensitivity (see table 2).

4.2.2 Reaction Time Data

Participants' reaction times were expected to increase from the 1- to 2-back condition. An ANOVA revealed that this was the case – there was a significant main effect of difficulty level (F[1,28]=74.99, p<.001). ANOVA also confirmed a main effect of task (F[3,26]=177.92, p<.001) – participants responded slowest to the SynBack task. This may have been affected, however, by the fact that SynBack had the longest stimuli lengths. Once again, no main effect of group (F[1,28]=2.16, p>.1) or interaction with group was found. A significant interaction of difficulty level and task type was found, however (F[3,26]=4.18, p<.05): participants showed the largest difference in reaction times between 1- and 2-back conditions during the SynBack task. Mean reactions for the younger and older unimpaired adults are reported in Fig. 1a and 1b, respectively; combined mean reaction times are reported in Fig. 1c.

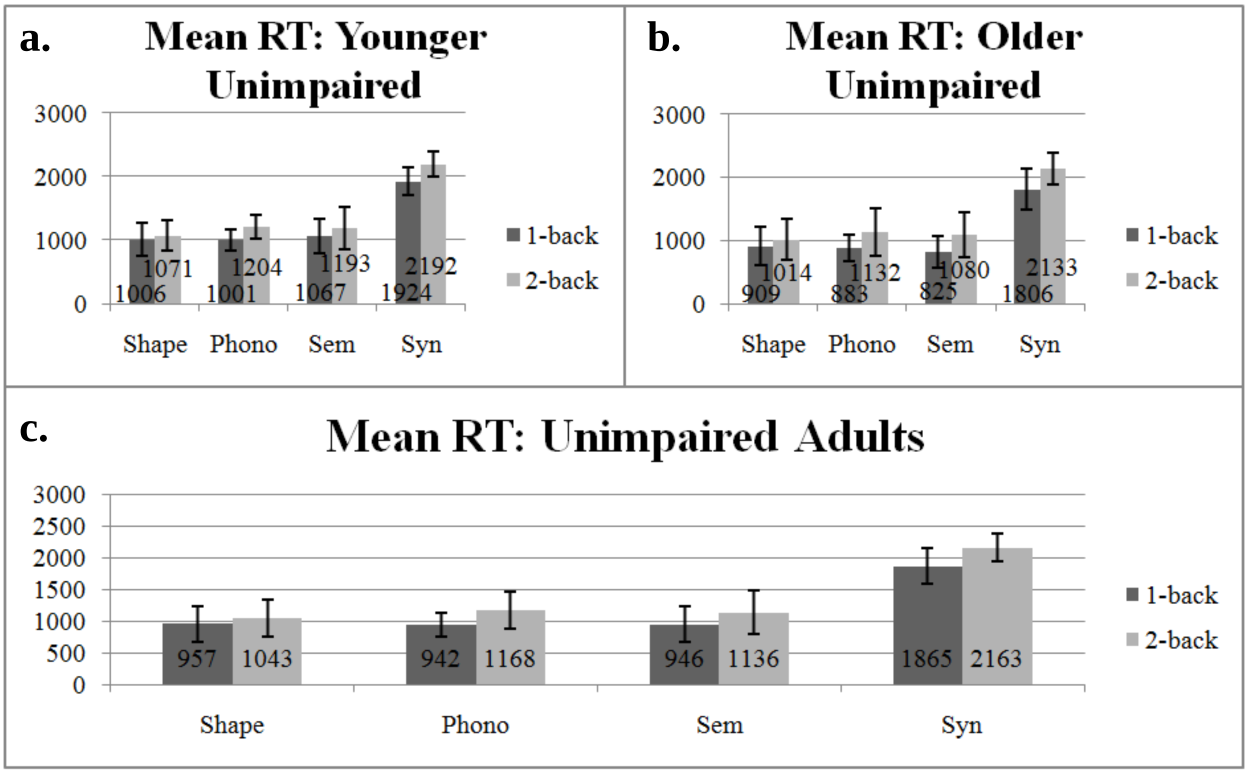

Figure 1: Mean Reaction Times of Older and Younger Unimpaired Adults

4.3 PWA

ANOVA revealed a main effect of task type on sensitivity scores (F[3, 6]=19.601, p<.01 [Greenhouse-Geisser DF: 1.9, 3.9]), and a marginally significant main effect of task type on accuracy scores (F[3, 6]=13.147, p=.052 [Greenhouse-Geisser DF: 1.19, 2.39]) and reaction times (F[3, 6]=12.825, p=.066 [Greenhouse-Geisser DF: 1.03, 2.07]). No significant interactions between task type and difficulty level were found. Although reaction time shows a numerical decrease from 1- to 2-Back on the SynBack task and most tasks show a decrease in accuracy and d' scores, paired-samples t-tests showed no significant differences between mean accuracy, reaction times, or d' scores in any task type. Individual accuracy and d' scores for each PWA are reported in Table 4, and reaction times are reported in Fig. 2.

Table 4: Accuracy & d' Scores of PWA

PWA 1								
	ShapeBack		PhonoBack		SemBack		SynBack	
	ACC	d’	ACC	d’	ACC	d’	ACC	d’
1-back	0.97	4.372	0.82	1.606	0.98	4.735	0.77	2.416
2-back	0.95	3.14	0.92	2.64	1	6.18	0.74	0.74
PWA 3								
	ShapeBack		PhonoBack		SemBack		SynBack	
	ACC	d’	ACC	d’	ACC	d’	ACC	d’
1-back	1	6.18	0.78	1.606	1	6.18	0.75	2.249
2-back	0.82	1.47	0.54	0.22	0.82	1.67	0.76	1.24
PWA 6								
	ShapeBack		PhonoBack		SemBack		SynBack	
	ACC	d’	ACC	d’	ACC	d’	ACC	d’
1-back	0.91	3.615	0.66	0.566	0.95	3.291	0.75	1.177
2-back	0.68	0.47	0.63	0.15	0.71	0.79	0.61	-2.17
Mean PWA								
	ShapeBack		PhonoBack		SemBack		SynBack	
	ACC	d’	ACC	d’	ACC	d’	ACC	d’
1-back	0.96	4.72233	0.75333	1.25933	0.97667	4.73533	0.75667	1.94733
2-back	0.81667	1.69333	0.69667	1.00333	0.84333	2.88	0.70333	-0.0533

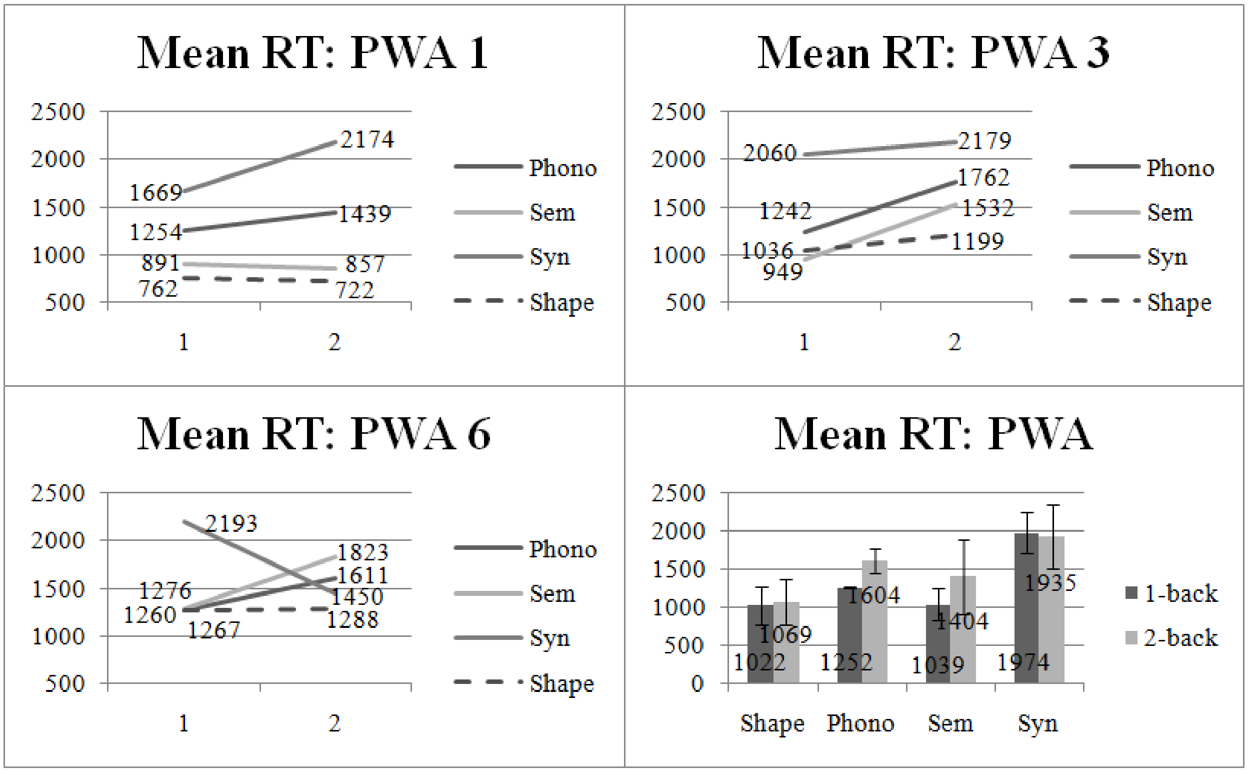

Figure 2: Reaction Times of PWA

Due to the small sample size of the PWA participant group, the data from the three members of that group were analyzed on an individual basis. Each PWA's D-Prime scores were compared to the ON subject group's. The comparisons were performed using Bayesian methods intended to estimate the difference between an individual's performance and that of a control group's (Crawford, Garthwaite & Porter, 2010).

PWA 1 did not show significant differences from the older cognitively normal group for any of the task types. PWA 3 showed a reliably lower performance on the PhonoBack task type in the 2-back condition ($p<.04$). PWA 3 also showed marginally worse scores in the 2-back condition for SemBack ($p<.06$) and SynBack ($p<.07$). Intriguingly, PWA 3 showed marginally better scores than the older controls for ShapeBack in the 1-back difficulty level ($p<.09$).

PWA 6 showed the most impaired performance in general. This participant performed reliably worse in the 1-back condition for SemBack ($p<.04$), and in the 2-back condition for all task types (PhonoBack: $p<.04$, SemBack: $p<.02$, SynBack: $p<.001$, ShapeBack: $p<.01$). Additionally, PWA 6 performed marginally worse than unimpaired controls on the 1-back condition of the PhonoBack task ($p<.06$).

All three PWA showed deficits on the syntactic PALPA subtest; Additionally, PWA 3 showed impairment on the semantic subtest, and PWA 6 showed reduced performance on the phonological subtest.

5.0 DISCUSSION

5.1 YOUNGER AND OLDER CONTROLS

Wright et al.'s original study reported only raw accuracy scores for target items, and showed significant decreases in performance for all tasks from 1-back to 2-back difficulty. The current study, however, found that raw accuracy scores *increased* for the SynBack task from the 1- to 2-back conditions for both the younger and older subject groups (fig. 1). This outcome suggested that simple accuracy scores are not the best measure of performance for n-back tasks, because a participant's tendency to respond or to abstain can affect measures of accuracy. A better measure for such tasks is a combination of Criterion scores, which measure how strong the stimulus needs to be before the subject will respond, and D Prime sensitivity scores, which measure how well the subject detects the presence or absence of the target stimulus. With these new measures, subjects showed no unexpected increases from the 1- to 2-back conditions.

After the original misleading measures were corrected, results fell into the expected patterns. SemBack was consistently the "easiest" of the tasks, as found in Wright et al.'s study, and SynBack was consistently the "hardest." These results indicate that the pattern of results shown by PWA in Wright et al.'s study was probably due to task difficulty, not information-specific deficits in WM. Reaction time showed a supra-additive effect for different tasks as the difficulty level changed. No significant differences between the older and younger groups were found.

The fact that most mean sensitivity scores for the Phono- and ShapeBack tasks do not significantly differ from one another is particularly interesting. One explanation for this relationship might be that the PhonoBack and ShapeBack tasks required both retention and suppression of predominantly sensory information – auditory in the case of PhonoBack and visual in the case of ShapeBack. An absence in variation between these two tasks may indicate that sensory information is processed similarly regardless of modality, in contradiction to Baddeley's model of a phonological loop and a separate visuo-spatial scratchpad. Detecting a correlation in performance between these two tests would support this interpretation. Alternately, the two tasks may simply be of similar difficulty, causing the similarity in performance. If this is the case, the tasks are not expected to show a strong correlation.

The results from the cognitively unimpaired participant group suggest that analyzing participants' performance on an n-back task using pure accuracy data may be misleading. As data collection from the PWA participant group progresses, a point of interest will be whether the same climb in accuracy from 1- to 2-back occurs. If so, previous studies using N-Back tasks to measure VWM in PWA will need to be re-evaluated. For example, Wright et al. (2007) provided no measures of d' sensitivity, only raw accuracy for target items ["hit rate"]. As no accuracy for non-target items is provided so that a false alarm rate may be determined, d' scores cannot be calculated to investigate whether the reported accuracy measures offer a realistic picture of participants' performance.

5.2 PWA

The performances of the PWA group seem to show no pattern, in contrast to the predicted results. PWA were expected to show type-specific deficits, and their performance on the n-back tasks were expected to correspond to the matched PALPA tasks.

PWA 1 shows no difference in performance from that of her cognitively unimpaired counterparts, suggesting that VWM reductions do not play a part in the difficulties aphasia causes. PWA 6, however, demonstrated universally poorer performance in the more demanding 2-back condition, suggesting a universal reduction in VWM. PWA 3 shows selective deficits in certain tasks, while demonstrating essentially normal performance in others; such performance suggests a sub-divided model of VWM, instead of one universal resource.

Additionally, the performance of the three PWA on the experimental task did not seem to correlate to performance on the PALPA. For example, while PWA 6 performed reliably worse than unimpaired controls on the SemBack task, she showed no deficit in the PALPA subtest that measures semantic understanding. By contrast, PWA 3 performed only marginally worse than unimpaired controls on SemBack but scored below normal limits on the PALPA semantic subtest. Two possible explanations exist for this pattern: either the n-back task tests a different cognitive process from that of the PALPA (e.g. the n-back task tests VWM while the PALPA tests processing skills, or the n-back task has an attentional component that the PALPA lacks), or the n-back tasks do not actually test the intended areas of language.

5.3 LIMITATIONS OF THE CURRENT STUDY

5.3.1 Theoretical Processes in the N-Back Task: Chen et al. (2008)

The n-back paradigm is not universally agreed to be a valid measure of WM, or at least not WM alone. The idea that a 2-back task somehow differs from a 0- or 1-back, as shown by its unique correlation to the TMT-A (Miller et al., 2009), is explored by Chen, Mitra, and Schlaghecken (2008). Interested in the mental subprocesses involved in completing an n-back task, the investigators used Electroencephalography (EEG) to measure neural response to 0-, 1-, and 2-back tasks. Tasks either focused on verbal information or spatial information; in the first experiment, stimuli were the same and subjects were asked to simply focus on one element, while in the second, stimuli were explicitly either verbal or spatial. Once again, stimuli were presented visually. Accuracy decreased with increasing N, was lower in the spatial task than the verbal, and was lower for match than for non-match stimuli.

Other dependent variables gathered from the EEG data were used to examine the processes involved in the n-back tasks. The authors hypothesized that a 0-back task, where all stimuli are compared to a pre-determined item, uses simply a "matching" mental subprocess, while a 1-back task involves both matching and "replacement," and a 2-back task involves these two subprocesses as well as a "'shift" subprocess. This hypothesis was supported; the neurophysiological activity associated with 'shifting' information appeared in the 2-back condition, but not in the 1- or 0-back. Furthermore, the study identified the 'replacement' subprocess as being associated with perceptual information, while the 'shift' subprocess is more connected with executive processing.

This study identifies a possible flaw of the n-back paradigm as a measure of VWM: while these tasks may test memory, they also test perception and processing. As demonstrated by Just and Carpenter (1992) and Caplan and Waters (1999), experts disagree on whether Verbal Working Memory is one single resource from which memory, perception, and executive processing all draw. The n-back paradigm may bias results toward this interpretation because it calls upon all three of these abilities during the 2-back level, meaning information types that draw on them will be affected. Experimenters who choose to use the n-back paradigm as a measure of WM should consider carefully about the assumptions about the nature of VWM that underlie this paradigm.

5.3.2 Sample Sizes

The insignificant effect sizes shown by the individual cognitively unimpaired groups, which become significant when the two groups' data are analyzed together, suggest that a sample size of 15 is too small for this test paradigm in unimpaired populations. Statistical power analysis of Wright et al. (2007) suggests, however, that a sample size of 15 should be enough to show significant effects in the language-impaired population. This means that effect sizes in the current study's younger and older control groups are smaller than those in Wright et al.'s PWA subjects. The current study's sample of 3 PWA is thus clearly insufficient.

5.3.3 SDT Analysis

While d' sensitivity scores seem to be more appropriate for use in analyzing data from the n-back paradigm, these scores do have certain drawbacks. Because d' is a parametric measure

(Stanislaw 1999), it is not well suited to the extremes of performance, i.e. hit rates and false alarm rates near one or zero. Additionally, there are various criterion measures and some disagreement as to which is the best (Stanislaw 1999). The criterion measure chosen for this study, c, was chosen because it indicates a participant's distance from neutrality, so that larger numbers are expected when a participant is more inclined to answer one way or another, and small numbers when he or she is not, allowing for easier identification of subject bias (Stanislaw 1999).

6.0 CONCLUSION

The definitive finding of this study is the misleading behavior of accuracy scores from the 1- to 2-back condition on the SynBack task. Given the nature of the n-back paradigm, wherein participants' false alarm rates can be as telling as their hit rates, analysis of performance must be more in-depth than a simple accuracy measure. The d' analysis performed in this study clearly shows that accuracy scores can indicate trends that do not exist due to participants' answer biases. This finding demands a change in the way future n-back tasks are scored, and in the way past n-back tasks are interpreted.

The original questions of this study, regarding the nature of VWM and its role in PWA's difficulty with sentence comprehension, are difficult to answer given the small size of the PWA participant group. There is no apparent pattern to the participants' performance – in fact, all three seemed to perform in fundamentally different ways. Given this conflicting evidence, and the fact that a statistical power analysis of Wright et al. (2007) indicates that a much larger sample is required to see reliable effects, the original research questions are perhaps better left unanswered until more data can be collected.

APPENDIX

SHAPEBACK TASK

This appendix will provide the five abstract non-nameable shapes (Hautzel et al., 2009) used in the ShapeBack task, along with the script used to present these shapes during the experiment. Each shape has been randomly assigned a corresponding number which the script uses to indicate the shape that would appear on the computer screen at that point. All of the information in this section was presented visually on a computer monitor; only one shape was presented at a time except in the instructions, where three shapes would appear at once to demonstrate the target pattern.

A.1 STIMULI

Table 5: Randomly Numbered ShapeBack Stimuli

1	2	3	4	5
				

A.2 PRESENTATION SCRIPTS

Target items are bold and underlined.

A.2.1 1-Back Condition

In this experiment, you're going to see a series of shapes. Try to remember what each shape looks like.

As you look at this series of shapes, there will be times when you'll see the same shape two times in a row. I want you to push the LEFT button (labeled "Yes" on keyboard) when the shape you just saw is the exact same shape as the one right before it.

An example of what you might see is:

4

(You wouldn't press the button)

2

(You wouldn’t press the button)

2

(Press the button!!)

The second and third shapes are the same, so you would push the button after seeing “2” the second time.

This will take less than 10 minutes.

Practice:	**3**	2	5	**1**
4	5	3	1	4
2	4	**3**	2	**4**
3	1	2	**2**	2
3	**1**	4	5	3
5	2	**4**	1	5
	5	2	3	**5**
Experimental:	1	3	**3**	2
3	**1**	2	4	1
1	2	**2**	**4**	**1**
1	5	4	2	2
4	4	2	5	5
2	**4**	4	**5**	**5**
2	3	**4**	4	
3	**3**	3	1	

A.2.2 2-Back Condition

In this part, you’ll be performing the same task, but this time push the button when the shape you just saw is exactly the same as the shape you saw two shapes ago. This means that there will be one shape in between the matching shapes.

An example of what you might see is:

1

(You wouldn't press the button)

5

(You wouldn't press the button)

1

(Press the button!!)

The first and third shapes are the same, so you would press the button after you see "1" the second time.

This task is more challenging than the last one, but just try to do your best.

This task will take less than 10 minutes.

Practice:	1	**3**	3	1
5	**4**	4	2	**2**
4	3	2	**3**	3
1	5	1	1	4
5	**3**	**2**	4	**3**
1	1	3	**1**	1
	5	1	2	4
Experimental:	2	**3**	5	5
4	3	2	4	**4**
5	**2**	5	**5**	3
3	5	**2**	2	2
5	3	2	4	**3**
2	**5**	4	3	1
3	1	1	1	4
2	1	**4**	**3**	2
5	3	2	4	**4**
4	2	5	2	

BIBLIOGRAPHY

Bayles, K., & Tomeoda, C. (1993). Arizona Battery for Communication Disorders of Dementia. Tucson, AZ: Canyonlands Publishing, Inc.

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. (1999). Verbal working memory and sentence comprehension. *Behavioral and Brain Sciences, 22*, 77-126.

Caramazza, A. & Zurif, E.B. (1976). Dissociation of algorithm and heuristic processes in language comprehension: Evidence from aphasia. *Brain and Language, 3*, 572-582.

Chen, Y., Mitra, S., & Schlaghecken, F. (2008). Sub-processes of working memory in the N-back task: An investigation using ERPs. *Clinical Neurophysiology*, *119*, 1546-1559.

Crawford, J. R., Garthwaite, P. H., & Porter, S. (2010). Point and interval estimates of effect sizes for the case-controls design in neuropsychology: Rationale, methods, implementations, and proposed reporting standards. *Cognitive Neuropsychology, 27*(3), 245-260.

Dickey, M. W. & Thompson, C. K. (2004). The resolution and recovery of filler-gap dependencies in aphasia: Evidence from on-line anomaly detection. *Brain and Language*, *88*, 108-127.

Folstein M.F., Folstein S.E., & McHugh, P.R.: " Mini-Mental State ": a practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res, 12, 189–198.

Friedmann, N., & Gvion, A. (2003). Sentence comprehension and working memory limitation in aphasia: A dissociation between semantic-syntactic and phonological reactivation. Brain and Language, 86, 23–39.

Gibson, E. (2000). The Dependency Locality Theory: A distance-based theory of linguistic complexity. In Miyashita, Y., Marantz, A. & O'Neil, W. (Eds.) Image, Language, Brain. MIT Press: Cambridge, MA. 95-126.

Hautzel, H., Mottaghy, F. M., Specht, K., Muller, H., Krause, B. J. (2009). Evidence of a modality-dependent role of the cerebellum in working memory? An fMRI study comparing verbal and abstract n-back tasks. *NeuroImage*, 47, 2073-2082.

Just, M. A. & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review, 99*(1), 122-149.

Kay, J., Lesser, R., & Coltheart, M. (1992). PALPA Psycholinguistic Assessments of Language Processing in Aphasia: Introduction. Lawrence Erlbaum Associates: East Sussex, England.

Kertesz, A. (1979). *Aphasia and Related Disorders*. New York: Grune and Stratton.

Kertesz, A. (2007). The Western Aphasia Battery-Revised. San Antonio, TX: Harcourt Assessment, Inc.

K.M., Price, C.C., Okun, M.S., Montijo, H., & Bowen, D. (2009). Is the N-Back task a valid neuropsychological measure for assessing working memory?. *Clinical Neuropsychology*, 24, 711-717.

Nation, K., Adams, J. & Snowling, M. (1999). Working memory deficits in poor comprehenders reflect underlying language impairments. *Journal of Experimental Child Psychology, 73*, 139-158.

Martin, R. C., Wu, D., Freedman, M., Jackson, E. F., & Lesch, M. (2003). An event-related fMRI Investigation of phonological versus semantic short-term memory. Journal of Neurolinguistics, 16(4–5), 341–360.

Miller, K.M., Price, C.C., Okun, M.S., Montijo, H., & Bowers, D. (2009). Is the N-Back Task a Valid Neuropsychological Measure for Assessing Working Memory?. Clinical Neuropsychology, 24, 711-717.

National Institute of Neurological Disorders and Stroke [NINDS]. (2009). Aphasia. *MedlinePlus.* Retrieved from <http://www.nlm.nih.gov/medlineplus/aphasia.html>.

Raven, J. (1956). Coloured Progressive Matrices: Sets A, Ab, B. London: H.K. Lewis and Co., Ltd.

Stanislaw, H. & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers*, 31(1), 137-149.

Wright, H. H., Downey, R. A., Gravier, M., Love, T. & Shapiro, L. P. (2007). Processing distinct linguistic information types in working memory in aphasia. *Aphasiology*, *21*, 802-813.