RICE UNIVERSITY

Separating Semantic and Phonological Short-term Memory in Aphasic Patients Using a Novel Concurrent Probe Paradigm

by

Heather Raye Dial

A THESIS SUBMITTED

IN PARTIAL FULFILLMENT OF THE

REQUIREMENTS FOR THE DEGREE

Master of Arts

APPROVED, THESIS COMMITTEE:

Randi Martin, Elma Schneider
Professor Psychology, Chair

Tatiana Schnur, Assistant Professor
Psychology

Jessica Logan, Assistant Professor
Psychology

HOUSTON, TEXAS

MAY 2014

ABSTRACT

Separating Semantic and Phonological Short-term Memory in Aphasic Patients Using a Novel Concurrent Probe Paradigm

by

Heather Raye Dial

Previous research suggests that short-term memory (STM) processes are separable into at least two buffers: a lexical-semantic and a phonological buffer. While there are multiple tasks used to measure phonological STM, only one task is commonly employed to test semantic STM, the category probe task. The current study used a novel paradigm, the concurrent probe paradigm (Shivde & Anderson, 2011), to measure semantic and phonological maintenance in aphasia patients. In Experiment 1, which evaluated semantic maintenance, we replicated the findings of Shivde and Anderson (2011) with older adults and revealed dissociations in patient performance depending on the type of STM deficit. The concurrent probe paradigm provided converging evidence with the category probe task in measuring semantic STM deficits. In Experiments 2 and 3, we applied the task to phonological maintenance. We replicated the findings of Shivde and Anderson (2011) with older adults, but for patients the results were less clear.

ACKNOWLEDGEMENTS

I would like to thank all of my committee members for their support and guidance. In particular, I would like to thank my advisor, Randi Martin, for her commitment to my success in this endeavor and her willingness to share her wealth of knowledge and experience. I would also like to thank my friends and family for their support and understanding during this process, without which my success would not be possible.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

Since its conception, researchers have been working diligently to define the construct of short-term memory (STM). While many researchers have agreed that STM processes are modality specific (e.g., Baddeley & Hitch, 1974; Martin, Shelton & Yaffee, 1994), there is still considerable debate as to how many separate storage systems may exist and what kind of information these storage systems may hold. For example, Baddeley and Hitch (1974) argued that short-term memory processes could be divided into two separate storage systems, a visual-spatial STM system and a phonological STM system; building off of this idea, Martin, Shelton and Yaffee (1994) included a phonological STM system, similar to the phonological loop proposed by Baddeley and Hitch (1974), as well as a semantic STM system in their model of STM and language. The current study focused on evaluating deficits in semantic and phonological STM in aphasic patients using a novel paradigm created by Shivde and Anderson (2011), which will be described in detail later in this paper. First, I will discuss some of the evidence suggesting that semantic and phonological STM are separate systems.

According to Martin and colleagues (Martin, Shelton & Yaffee, 1994; Martin & He, 2004), STM in the context of language is separated into at least two different types: semantic STM and phonological STM. Semantic STM is involved in maintaining word meanings in a highly accessible state whereas phonological STM is involved in maintaining phonological representations for words and non-words. During language comprehension, semantic STM is argued to be involved in maintaining the meanings of several words in a phrase prior to their integration (e.g., in maintaining the meanings of prenominal adjectives in "short blonde hair" prior to their integration with "hair") (Martin & Romani, 1994; Martin & He, 2004). Semantic STM has also been argued to be involved in maintaining the meanings of words in a phrase prior to their articulation (e.g., in producing "rusty old wagon") (Martin & Freedman, 2001; Martin, Miller, & Vu, 2004; Martin, Crowther, et al., 2010). In contrast, phonological STM appears to play little role in comprehension, but is important during verbatim sentence repetition (Martin et al., 1994) and when repeating non-words or words that have no associated semantic context, such as in repeating words from foreign languages (Hulme, Maughan & Brown, 1991). Although there is evidence that supports this contention of a separation of semantic and phonological capacities, there is still considerable debate regarding the role of STM in language processes. For example, Just and Carpenter (1992) argued that there is a single STM capacity involved in language comprehension; others have argued that there are multiple STM capacities that are associated with specific processes (Barnard, 1985; Caplan & Waters, 1999; Martin, Shelton & Yaffee, 1994). The remainder of this paper will focus on short-term memory processes and will contain a more in-depth discussion of the current research on separable short-term memory systems.

Evidence for the distinction between semantic and phonological STM comes from neurally healthy and impaired populations. For example, unimpaired subjects generally exhibit a greater span for words than non-words (Crowder, 1978; Hulme, Maugham & Brown). This suggests that the semantic information associated with words increases the number of words that can be retained as compared to non-words, even though the phonological representations may be equivalent in length. While the semantic information for word meanings is stored in long-term memory (LTM) these meanings are activated and stored in semantic STM (e.g., during speech comprehension). For example, in sentence comprehension, subjects must use semantic STM to maintain words within a phrase before integrating them with the rest of the sentence (e.g., "Rocks, trees and shrubs grew in the yard"; Martin & He, 2004). In addition, the short-term retention of foreign language words increases when semantic knowledge is associated with the words instead of simply memorizing their phonology (Hulme, Maughan & Brown, 1991). If semantic and phonological STM were operating via the same mechanism, short-term list recall for foreign language words should not be affected by the availability of the meanings of the words.

Other evidence for semantic STM storage comes from aphasic patients who exhibit a double dissociation between phonological and semantic STM impairments (Martin, Shelton & Yaffee, 1994; Martin & He, 2004). Martin et al. (1994) tested two patients with reduced STM spans but who showed different patterns of impairment on STM tasks. Patient EA exhibited symptoms of a phonological impairment: she did not display phonological similarity effects for visually presented words (i.e. decreased ability to recall a list of words when they are phonologically similar) but still showed an advantage for words over non-words, indicating intact semantic STM. Patient AB presented the opposite pattern. Both individuals completed a rhyme probe task and a category probe task. In probe tasks, subjects are presented with a list of words followed by a short pause and a “probe” word. In the category probe task, the subject’s task was to determine whether the probe word was in the same category as any of the previously presented list words; in the rhyme probe task, subjects determined whether the probe word rhymed with any of the previously presented list words. For example, in the category probe a subject may hear a list of words, “bear, daisy, broiler, hurricane” and then hear the word “monkey.” In this case the probe word was a member of the same category as bear (i.e. animals) and so the correct response would be yes. The rhyme probe task works in a similar fashion, with the exception that subjects judged whether the probe word rhymed with a list word (or not). EA performed better than AB on the category probe task whereas the reverse was true for the rhyme probe task. These results indicated that EA and AB had separable patterns of STM deficits, namely that EA suffered from a phonological STM deficit and AB suffered from a semantic STM deficit. Martin and He (2004) extended these findings to patient ML, an aphasic individual who exhibited a stronger dissociation than did AB between semantic knowledge and semantic STM capacities. ML demonstrated intact semantic knowledge on general semantic tasks such as picture naming and single word comprehension, but had impaired semantic STM capacities as tapped by tasks such as the category probe. This dissociation indicated that STM capacities are separate from lexical knowledge.

Martin, Lesch and Bartha (1999) further distinguished between input and output phonological STM. MS, who was anomic, exhibited an output phonological deficit. For example, in an examination of errors in spoken list recall, MS tended to make errors by producing words that were phonologically similar to words in the list, suggesting that MS was able to retain phonological information from words in the list. Interestingly, he was often able to give the meaning of the words that he could not produce. Within this list recall paradigm, the researchers had MS attempt to recall words that he could or could not name in a picture-naming task and compared this to list recall on words that he could or could not comprehend. MS was worse at list recall for words he could not name than words he could name, but there was no difference in performance for words on which he made errors on the comprehension tests. His accurate descriptions of the words he could not recall suggested that phonological information was appropriately activating the semantic information for the words, demonstrating that his input phonological STM was intact. Further, MS performed near ceiling and within or above the range of controls on tasks that tapped input phonological STM, such as the rhyme probe task. It was thus hypothesized that he was anomic due to an inability to activate phonological representations. In other words, patient MS suffered from a deficit in activating output phonological representations and thus these representations could not be connected to an output buffer. This deficit contrasts that of patient EA. EA had normal spontaneous speech production, but her performance was impaired on memory probe tasks (which do not require list output); these findings suggested that EA had an input phonological STM deficit. This double dissociation provided evidence for a separation of input and output phonological STM. Martin, Lesch and Bartha (1999) thus proposed a model of the role of STM in language comprehension and production. This model specified separate buffers for the phonological and semantic STM components, with the phonological component being broken down further into separate input and output buffers.

Following the publication of the Martin, Lesch and Bartha (1999) model of STM, several researchers began to investigate semantic and phonological STM processes. Majerus, Van der Linden and Renard (2001) tested the model on a group of stroke patients with left hemisphere damage. The patients were administered both the rhyme and category probe tasks, argued to tap phonological and semantic STM, respectively; then, patients completed other phonological STM tasks (i.e., consonant-vowel (CV) and consonant-vowel-consonant (CVC) span and word and non-word list repetition). All of the left hemisphere damaged patients were impaired on these tasks relative to controls, but the pattern of deficits among the patients was quite variable. First, the patients' performance replicated the findings of Martin and colleagues (e.g. Martin, Shelton & Yaffee, 1994; Martin, Lesch & Bartha, 1999) in that some patients were relatively more impaired on the category than the rhyme probe task and some patients were relatively more impaired on the rhyme than category probe task. Importantly, the category and rhyme probe tasks did not correlate with each other, suggesting they measure different underlying constructs. Additionally, performance on the rhyme probe task was correlated with performance on the CV and CVC span tasks as well as the word and non-word repetition tasks, but category probe performance was not correlated with any of these measures. Thus, Majerus, Van der Linden and Renard (2001) replicated the dissociable pattern of STM deficits of Martin and colleagues and assessed the validity of the category and rhyme probe tasks as measures of semantic and phonological STM, respectively.

Haarman and Usher (2001) also investigated separable STM processes. Haarman and Usher (2001) focused their research on semantic STM. They had normal subjects perform an immediate free recall test. They also had subjects complete a LTM version of this task by inserting a distractor task in between list presentation and recall. The authors wanted to determine whether semantic STM processes actually exist or whether semantic LTM influences performance on STM tasks. The critical factor in this experiment was how the lists were arranged. Each 12-item list was set up so that pairs of semantically related words were right next to each other or separated by a distance of 5 words. Glanzer (1969) had previously shown that adjacent presentation of semantically related words facilitated recall of lists. He attributed this effect to both STM processes, whereby related words support each other in STM, and to transfer of information to episodic LTM. Haarman and Usher (2001) sought to determine the unique contributions of the hypothesized STM and LTM processes that led to the adjacency results of Glanzer (1969). Thus, the researchers asked, if we factor out performance on a LTM version of the task, will STM effects still exist? They found that the effect of separation was limited to the recency portion of the free recall curve, which is traditionally associated with STM. Importantly, the researchers found that individuals were able to correctly recall more items when semantically related words were adjacent rather than separated, even after factoring out LTM performance on this task. The authors concluded that semantic STM was responsible for the effect and that semantic STM is capacity limited, as evidenced by the benefit of semantic relatedness in recent, but not earlier, positions.

Researchers have also examined the neurological basis of semantic and phonological STM systems. In an fMRI study, Shivde and Thompson-Schill (2004) had subjects complete a delayed synonymy judgment task (semantic condition) and a delayed phonological judgment task (phonological condition); the researchers were interested in whether the semantic and phonological conditions would elicit activation in different regions of healthy subjects' brains. In the semantic condition, subjects were first presented with a word for 2 seconds that they were told to remember by focusing on its meaning. Following a 10 second delay, a second word appeared on the screen. Subjects were required to make a decision as to whether the second word was synonymous with the first word. The phonological condition had the same basic arrangement. First, subjects were presented with a word for 2 seconds and told to remember it by focusing on its sound. Following a 10 second delay, a non-word appeared on the screen. Subjects were required to make a decision as to whether the non-word shared a vowel sound with the maintenance word. The results showed that the semantic condition elicited greater activation in bilateral inferior frontal gyri and the left middle temporal gyrus, while the phonological condition elicited greater activation in the left superior parietal lobe. These data suggested that semantic and phonological maintenance processes rely on different brain structures and thus are separable processes.

Hamilton, Martin and Burton (2010) also used fMRI to investigate the neural basis for semantic STM. In their task, subjects were required to detect semantic anomalies. In an acceptable phrase, subjects were presented with a series of words, for example, "green, shining, bright emerald." An anomalous version of this phrase was "green, shining, bright sun." In this condition where the adjectives are listed before the noun (i.e., the before condition), subjects are required to maintain prenominal adjectives in a highly accessible state (i.e., semantic STM) until they can be integrated with the noun. In a second condition, the adjectives appeared after the noun in a phrase such as "emerald, big, green, shining." In this condition (i.e., the after trial type) subjects can integrate the adjectives with the noun as they appear rather than being required to maintain all adjectives until the noun appears. Thus, the after condition places less demand on semantic STM than does the before condition. The results showed that there was greater activation in left inferior frontal gyrus and middle temporal gyrus in the before than in the after conditions. However, no greater activation was seen for the adjectives before vs. adjectives after in regions thought to support phonological STM. These findings suggest that semantic STM processes engender distinct patterns of brain activation compared to phonological STM processes.

A recent study by Shivde and Anderson (2011) further investigated semantic STM using a novel paradigm. The authors argued that their new paradigm escaped certain flaws associated with other paradigms, namely the possibility that long-term semantic memory or long-term semantic priming caused the reported effects. For example, in regards to the Haarman and Usher (2001) findings, Shivde and Anderson (2011) argued that the observed "semantic recency" effect could also be described using a long-term semantic priming explanation. In this context, the presentation of the first word in a semantically related word pair primes other semantically related words in LTM via spreading activation (e.g., Collins & Loftus, 1975). This boost in activation for semantically related words could spread to the phonological forms of the words, causing them to be in a highly accessible state, leading to easier encoding of the second word in the word pair, and perhaps a stronger phonological representation or more time for rehearsal. Thus, according to Shivde and Anderson (2011), LT semantic priming, not “semantic recency”, might be a better term for this effect.

In order to avoid the possible confounds associated with other STM paradigms, Shivde and Anderson (2011) created a novel paradigm, called the concurrent probe paradigm. Previous studies had used immediate recall as their measure of semantic STM, but, as mentioned previously, the authors argued that other studies of semantic STM could not rule out episodic or semantic LTM as the mechanism by which semantic effects arose. Thus, Shivde and Anderson (2011) developed the concurrent probe paradigm, a task in which semantic maintenance is measured during the maintenance interval instead of during recall. The authors argued that the state of items currently being maintained should differ from items that are not currently being maintained; that is, items being currently maintained should be in a state of heightened accessibility relative to those items not being currently maintained. Accordingly, they manipulated several factors to attempt to determine whether items were in a state of active maintenance (or not). In this paradigm, subjects were first presented with a word printed in red (the maintenance word) and instructed to remember the word by focusing on its meaning; this maintenance word was maintained for varying lengths of delay (the maintenance interval). After presentation of the maintenance word, subjects were required to make a series of lexical decisions to items in black font during and after the maintenance interval. At some point during the series of lexical decision, a blue target item appeared. When it appeared, subjects were required to judge whether it meant the same thing as the red word. Additionally and unbeknownst to the subject, the seventh item in the sequence of lexical decisions plus target item presentation was always a probe word. The probe word could be semantically related, phonologically related, or unrelated to the maintenance word. As the subject was not told about the existence of the probe item, they presumably treated it as if it were any other lexical decision trial. The semantically related probes were always weak associates (e.g., maintenance item: go, probe item: race) in order not to draw attention to the relation of the maintenance and probe item. Note that the probe word could come either during the maintenance interval (i.e. between the maintenance and target word; *during* trial type) or after the maintenance interval (i.e. after the target word; *after* trial type). Figure 1 is a schematic diagram of the semantic maintenance experiment from Shivde and Anderson (2011).

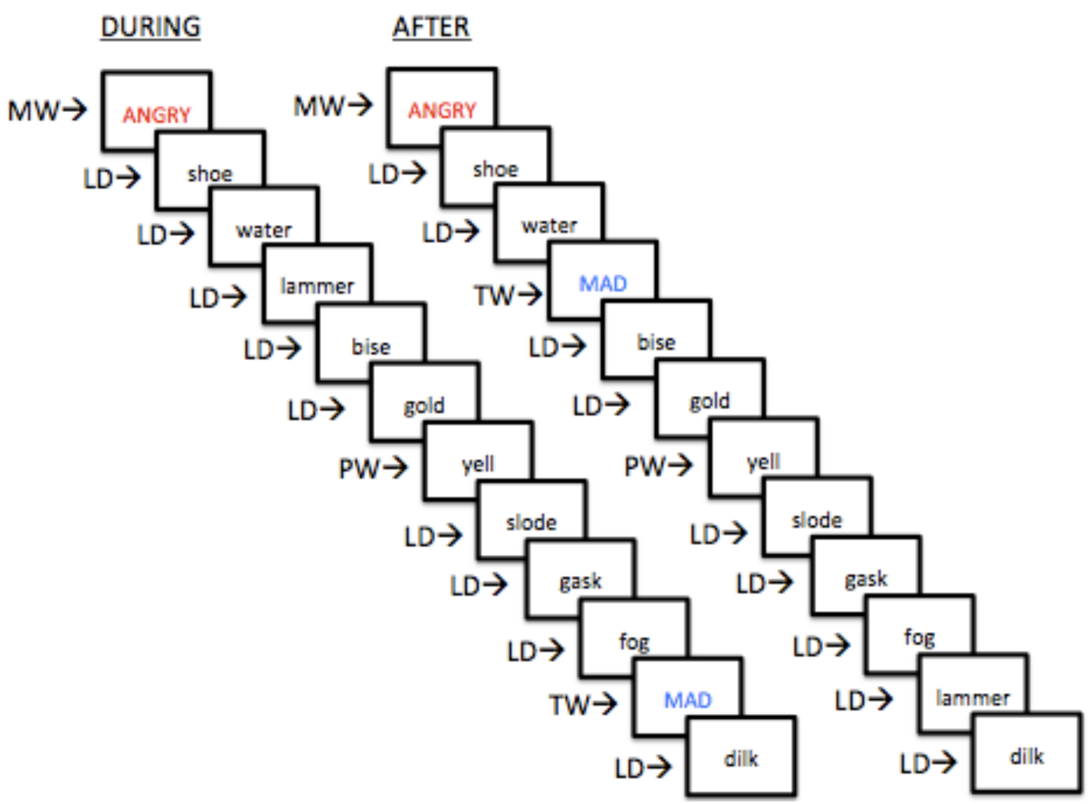

Figure 1. Schematic diagram of the concurrent probe paradigm (Shivde & Anderson, 2011). MW=maintenance word, LD=lexical decision, PW=probe word, TW=target word.

To contrast the aforementioned semantic maintenance condition, Shivde and Anderson (2011) also had subjects complete a phonological version of the task. In the phonological version, subjects were told to focus on the sound of the maintenance word to retain it during the maintenance interval. Additionally, instead of making a synonymy judgment to the target item, subjects were required to determine whether the maintenance word shared a vowel sound with the target, in this case a non-word. For example, happy and zee share a vowel sound, so this response would be yes.

In the semantic maintenance condition, Shivde and Anderson (2011) examined whether there was a difference in lexical decision response times between semantically related and unrelated probes and whether there would be an interaction with trial type (i.e., whether the probe word was during or after the maintenance interval). They argued that if subjects were actively maintaining the maintenance word using semantic STM, then response times for a lexical decision for a semantically related probe word should be slower than unrelated probes. Additionally, they argued that this difference should exist only in the *during*, but not *after*, trial type; that is, during maintenance, one would expect a semantic relatedness effect, but after maintenance has ended, one would not expect to see a difference between semantically related and unrelated probe lexical decision response times. Hypothetically, when the subject is actively maintaining the red maintenance word via semantics, activation automatically spreads to semantically related concepts. In order to focus on the actual maintenance word, Shivde and Anderson proposed that subjects must inhibit the other semantically related words; thus, when a lexical decision trial containing a semantically related word appeared, subjects should be slower to make a response because the word is in a lowered state of activation compared to unrelated words. In contrast, because subjects are maintaining the meaning of the word and not its phonological form, they should show no difference in response times to phonologically related words relative to unrelated words. Their hypotheses were supported by the data. In the semantic maintenance condition, subjects' response times for semantically related probes words were significantly greater than response times for unrelated probe words in the *during*, but not the *after*, trial type. In contrast, there was no difference in response times for phonologically related compared to unrelated words in the *during* or *after* trial types. In the phonological maintenance condition subjects took significantly longer to make a lexical decision for a phonologically related probe word compared to an unrelated probe word in the *during*, but not the *after*, trial type (i.e., phonological relatedness effect), but showed no difference in response times for semantically related compared to unrelated items in *during* or *after* trial types. This provided evidence of a double dissociation between semantic and phonological maintenance. The authors argued that this paradigm provided direct evidence for a dissociation between semantic and phonological maintenance in a way that avoided the confounds of other STM experiments.

The purpose of the current study was to evaluate semantic and phonological STM in older adults and aphasic patients using the concurrent probe paradigm (Shivde & Anderson, 2011) and to provide evidence that the concurrent probe paradigm is a valid measure for these types of STM, especially semantic STM. While there are multiple measures routinely used for phonological maintenance (e.g., digit span, digit matching, rhyme probe), there is only one measure routinely used for semantic maintenance - the category probe task. In previous research, researchers have sometimes had subjects perform a three-choice synonymy judgment task to measure semantic STM (Allen, Martin & Martin, 2012; Freedman & Martin, 2001). In this task, three words are spoken aloud and also shown visually, printed diagonally across a sheet of paper and subjects must determine which two of the three words are synonyms. Despite the use of this task to measure semantic STM by some researchers, potential confounds arise when using this task with patients. For example, the task is typically untimed and given the lack of a timing deadline, some patients may take a very long time to complete the task and thereby manage to perform at a very high level, despite severely impaired semantic STM. Second, the task taps a fine degree of semantic knowledge as the distractor items are typically highly related. Thus, poor performance could be due to a disruption of precise semantic knowledge and consequently, a comparison with a two-choice version of the task with fewer memory demands is sometimes employed to control for the semantic knowledge required (Freedman & Martin, 2001). Given the problems associated with the three-choice synonymy judgment and the relative disparity in the number of semantic compared to phonological STM tasks, the current study sought to provide evidence that the concurrent probe task is useful in measuring semantic STM, thus establishing another measure for semantic STM. To do this, we tested older adults and aphasic patients on various versions of the concurrent probe task. It is important to note that we were not interested in age-related decline on STM tasks. Rather, the purpose of testing age-matched controls was to provide a baseline for evaluating patient performance. Of course, if the findings for older adults did not replicate those of Shivde and Anderson (2011), then it would be uninformative to test patients, as no strong conclusions could be made.

In Experiment 1, modeled after the semantic maintenance condition of Shivde and Anderson (2011), we evaluated whether semantic and phonological STM patients would exhibit a semantic relatedness effect in *during* but not *after* trial types. First, we tested neurally healthy older adults to evaluate if they would exhibit this pattern. If neurally healthy older adults were able to use semantic maintenance strategies in the concurrent probe paradigm, and thereby exhibit the semantic relatedness effect, then we could assume that these processes are generally spared with aging. Having shown this effect in older adults, we then proceeded to test aphasic patients. We tested patients with dissociable STM deficits like those of Martin and colleagues (e.g., Martin, Shelton & Yaffee, 1994; Martin & He, 2004), namely aphasic patients with semantic or phonological STM deficits. A detailed description of the patients is contained in the Method section.

For our semantic STM patients, different possible outcomes were predicted depending on the source of their STM deficit. Based on previous research, two hypotheses have been put forward concerning the source of deficits in the short-term retention of semantic information: difficulty inhibiting irrelevant verbal information (Hamilton & Martin, 2005) and overly rapid decay (Martin & Lesch, 1996). If semantic STM deficits are caused by deficits in inhibiting semantically related but irrelevant information, then patients might show shorter response times for semantically related compared to unrelated probe words. For example, Hamilton and Martin (2005, 2007) showed that patient ML, a classic semantic STM patient, had an inability to inhibit irrelevant stimuli in a "recent negatives" probe task and this contributed to his semantic STM problems. Thus, if subjects have difficulty with inhibitory processes, maintaining a word via semantic maintenance would spread activation to semantically related words and patients would have difficulty inhibiting them. Then, when a semantically related word appeared in the lexical decision trial, they would be primed for this word and have decreased response times relative to unrelated words, but this would only be expected while subjects are actively maintaining the maintenance word, which occurs in the *during* trial type. In the *after* trial type, we would not expect any difference between semantically related and unrelated words.

On the other hand, if patients' semantic STM deficits are caused by overly rapid decay of semantic representations, we would expect patients' response times to be equivalent for semantically related and unrelated probe words in *during* and *after* trial types. It is possible that the patients' level of activation for the maintenance item will be less than that of healthy older adults; thus, activation may not spread to other semantically related words and there would be no need to inhibit those words in memory. In this case, we would expect response times of semantically related words to be equal to those of unrelated probe words in *during* trial types. Further, since the probe word appears after the maintenance interval has ended in the *after* trial types, we would not expect any difference between semantically related and unrelated probe words in the *after* trial type. Given the prediction that patients may be unable to maintain items in a heightened state of activation, one might assume that patients would exhibit increased error rates or be unable to complete the task at all; however, our selection criteria for patient participation in the task included that any patient tested in the current experiments have spans that are greater than or equal to one item. Thus, patients were selected based on the ability to maintain at least one word in memory at a time, as is the case for the current task (i.e., maintaining the maintenance word), though such a span would not guarantee that they cold maintain the word while performing the lexical decision task.

If patients failed to show a semantic relatedness effect in the *during* trial type, it is possible that they instead relied on a phonological code to retain the maintenance word. In this case, response times for semantically related and unrelated probe words would be equal. Thus, in order to draw any conclusions as to a lack of semantic relatedness effects in patients, we examined the possible use of phonological maintenance strategies when instructed to use semantic maintenance. To do so, neurally impaired and unimpaired participants took part in Experiment 2, in which subjects completed the same type of task as Experiment 1, but with different materials. In Experiment 2, subjects were instructed to use semantic maintenance, but, crucially, the probe words of interest were phonologically related rather than semantically related to the maintenance word. If control subjects or patients were using phonological maintenance strategies to aid performance, then we expected to see a phonological relatedness effect, whereby response times for phonologically related probe items are longer than for unrelated probe items in *during*, but not *after*, trial types.

If our phonological STM patients have relatively spared semantic STM, then we expected them to show a semantic relatedness effect in Experiment 1. Even if they cannot remember the phonological form of the maintenance word, they should be able to determine whether the target item was synonymous with the maintenance item because the semantic representation of the item would be maintained in a heightened state of activation. In addition, we would not expect our phonological STM patients to show a phonological relatedness effect in Experiment 2. If they have deficits in phonological STM, then it is unlikely that they would adopt a phonological maintenance strategy when instructed to use a semantic maintenance strategy, as this might make the task more difficult for them. However, this cannot be ruled out entirely. Due to the nature of the task, only one word is required to be maintained over the maintenance interval. Given the minimal demands that this places on STM, it is possible that the phonological STM patients could constantly refresh their memory by repeating the word across the maintenance interval. If this is the case, the same predictions apply to phonological STM patients for Experiment 2 as for the semantic STM patients in Experiment 1. Namely, if they have deficits in phonological STM due to an inability to inhibit related words, they may show priming (i.e., decreased response times) for phonologically related compared to unrelated probe words in the *during*, but not *after*, trial types. In contrast, if their deficit is due to a failure to activate a word to the same degree as healthy control subjects, they may show no difference between phonologically related and unrelated words in the *during* or *after* trial types.

Experiment 3 further evaluated phonological maintenance strategies in a phonological version of the concurrent probe paradigm. In Experiment 3, we had subjects maintain the maintenance word via phonological maintenance and make a judgment on a target nonword as to whether it was pronounced the same way as the maintenance words (e.g., for the maintenance word “brain,” the target was the pseudohomophone “brane”). Additionally, the critical probe words were phonologically related, as opposed to semantically related, to the maintenance word. For our phonological STM patients, we expected that they would either show no difference in the amount of time it takes to respond to phonologically related compared to unrelated probe words (i.e., if they have difficulty keeping an item in a heightened state of activation via phonology) or that they would be primed for phonologically related probe words (i.e., if they had difficulty inhibiting phonologically related words). On the other hand, we expected our semantic STM patients to show a phonological relatedness effect.

Thus, across Experiments 1-3, we sought to further establish a dissociation between semantic and phonological maintenance. Additionally, we were interested in establishing the usefulness of the various versions of the concurrent probe paradigm in evaluating semantic and phonological STM in stroke patients. So, we compared our patient’s performance on the 3 versions of the concurrent probe task with their performance on various measures of STM. Specifically, we evaluated whether patient performance on the semantic version of the concurrent probe task used in Experiment 1 and the phonological version used in Experiment 3 were related to patient performance on tests tapping STM for semantics and phonology (i.e., the category and rhyme probe tasks, described in the following section, Screening assessments). We expected that performance on Experiment 1 would be related to performance on tests tapping STM for semantic information, and performance on Experiment 3 would be related to performance on tests tapping STM for phonological. In this way, we were able to evaluate the effectiveness of the concurrent probe task as a new measure that can be used to evaluate STM deficits in patients. The screening assessments and STM assessments are described in the following section.

Screening Assessments

The following screening assessments were used as criteria for selecting patients for the current study. The rationale and importance of using each of these screening measures is included within the description of each task. In general, these screening assessments were used to rule out potentially confounding deficits (i.e., semantic knowledge, single word processing) and to ensure that patients would be able to perform the concurrent probe task. Data for most of the screening assessments were acquired from the patient database of the Rice University Brain and Language Lab. All patient data had been collected on these tasks before the current study was undertaken, with two exceptions. First, DW's original score on the Pyramids and Palm Trees Test (PPT) was 83%, but this testing took place over 5 years prior to the current study. Thus, DW was retested on the PPT. Additionally, PP had never been tested on the visual lexical decision task prior to the current study and so he was tested on this task after one of the experimental sessions. A summary of the screening tests is contained in Table 1. A summary of the patients is presented in Table 2 and patient performance on the various screening tests listed in Table 1 is presented in Table 3.

Table 1

Screening Assessments and Experimental Tasks

Task and Description	
Screening Assessments	
Synonymy Judgment	Subject must determine which 2 of 3 diagonally presented words are synonymous.
Pseudohomophone Judgment	Subject must determine whether a word and a non-word would be pronounced the exact same way.
Category Probe	Subject must determine whether a probe word is in the same category as any item within a previously presented list of words. This task is auditorily presented.
Rhyme Probe	Subject must determine whether a probe word rhymes with any item within a previously presented list of words. This task is auditorily presented.
Visual Lexical Decision	Subject must determine whether a visually presented item is a word or a non-word.
Pyramids and Palm Trees (PPT)	Subject must determine which 1 out of 2 pictures is most closely related to a third picture.
Single-word picture matching (PWM)	Picture displayed to subject and experimenter asks "Is this a _______?" Subject must determine whether the picture is the item.
Experimental Tasks	
Experiment 1A	Concurrent probe task with semantic maintenance and semantically related probe word
Experiment 1B	Concurrent probe task with semantic maintenance task and semantically related probe word; 437 ms added between each lexical decision trial
Experiment 1C	Concurrent probe task with semantic maintenance task and semantically related probe word; words that were in the *during* trial type in Experiment 1A are now in the *after* trial type, and vice versa
Experiment 2	Concurrent probe task with semantic maintenance task and phonologically related probe word
Experiment 3	Concurrent probe task with phonological maintenance task and phonologically related probe word

Table 2

Patient Background Information

Patient	Age	Years post-stroke	Education	Damage
DW	57	14	16	Left frontal including BA44 and 45, with extension into middle frontal gyrus; insular damage also present
LC	66	10	11	Left parietal; additional small subcortical infarcts of the posterior and lateral right parietal lobe
PP	67	2	20	Left superior temporal gyrus, insula
RI	86	11	12	Left temporal lobe and portions of the left posterior parietal lobe

Table 3

Patient and Control Data for Screening Assessments

Patient	PWM % correct	Lexical decision % correct	PPT % correct	Synonymy judgment % correct	Pseudohomophone judgment % correct	Category probe span estimate	Rhyme probe span estimate
DW	95	95	92	88	96	2	4
LC	98	98	100	96	98	2.71	2
PP	99.5	99	98	96	98	3.5	4
RI	98	94	98	100	99	3	1.67
Controls	---	99	98	95	---	4-7	3.84-9
		SD=.52	(94-100)	SD=5.2			

Note. With exceptions following, control data are referenced in the task description. Controls were not tested on the PWM task because Martin et al. (1999) assumed control participants would obtain 100% correct. Controls were not tested on the pseudohomophone judgment task for the same reason. Synonymy judgment, category probe and rhyme probe measures were all obtained from the control subjects tested in Experiments 1-3. Category and rhyme probe data presented represent the range of controls tested for the current experiment.

Pyramids and Palm Trees

The Pyramids and Palm Trees test (PPT) is a published test that was created by Howard and Patterson (1992) to assess semantic knowledge in patients with aphasia. The picture subtest was used as a qualifying measure for patients to participate in the current study. In this test, subjects are presented with 3 pictures, with one of those pictures placed above the other two.

Subjects are instructed to point to which of the bottom two pictures is most closely related to the top picture. The dependent measure reported in Table 3 is accuracy. The PPT was an important screening assessment for the current study because we wanted to be sure that our patients did not have deficits in semantic knowledge that would confound the STM findings.

Single Picture-Word Matching

In the single picture-word matching task (PWM), patients are presented with a picture and asked "Is this a ________?" Patients indicate whether or not the presented picture is that which was spoken by saying yes or no. There are four conditions in this task: one in which the target picture is correct, one in which the picture is a semantically related foil, one in which the picture is a phonologically related foil, and one in which the picture is an unrelated foil. The dependent measure reported in Table 3 is overall accuracy averaged across the four conditions. This task was used as a screening measure because it measures single word processing. Again, we did not want patients who had deficits in single word processing to be tested in the current study due to the possible confound with deficits of semantic STM.

Visual Lexical Decision

The visual lexical decision task was subtest 25 of the PALPA test (Kay, Coltheart & Lesser, 1992). The task is administered via pen and paper and the subject must determine whether each item is a word or not. They simply put a check mark next to the real words. The dependent measure reported in Table 3 is overall accuracy. We used visual lexical decision as a screening assessment because patients needed to be able to perform visual lexical decision tasks in order to do the concurrent probe task. Patients were only selected if they were able to perform this task at 90% or greater accuracy.

Additionally, two other tests were administered to the patients to determine their ability to perform synonymy and pseudohomophone judgments without the concurrent lexical decision task. If patients were unable to perform these tasks individually, then they would be unable to perform these tasks in the concurrent probe paradigm. These tests are described below.

Synonymy Judgment Task

In this task, subjects are presented with 3 words that are printed diagonally across a sheet of paper. The subject has to select the two items that are synonymous with each other. The dependent measure reported in Table 3 is accuracy.

Pseudohomophone Judgment Task

In this task, subjects were presented with the stimuli used in Experiment 3. The subject was presented with one word (the red maintenance item in Experiment 3) immediately followed by a non-word (the blue target item in Experiment 3). The subject then responded whether the two items would be pronounced the same way by making a button press. This task was administered using Psyscope (Cohen, MacWhinney, Flatt & Provost, 1993) on the same apparatus as Experiments 1-3. The dependent measure reported in Table 3 is accuracy.

Short-Term Memory Assessments

Category Probe

In the category probe task, the subject's task is to determine whether a probe word is in the same category as any of a previously presented list of words. For example, a subject may hear a list of words, "bear, ant, broiler, hurricane" and then hear the word "monkey." In this case the probe word was a member of the same category as bear (i.e. animals) and so the correct response would be yes. In this task, subjects begin at list lengths of 1 and subsequently progress to list lengths of 2, 3, 4, etc (the example above is that of list length 3) until their performance drops to or below 75% accuracy. A subject's category probe span is determined to be the estimated list length at which they perform at 75% accuracy. If, for example, a subject is 80% accurate on list length 2 and 70% accurate on list length 3, their span is determined via linear interpolation to be 2.5.

Rhyme Probe

The rhyme probe task is similar to the category probe task, with the exception that the probe word rhymes with a list word (or not) in place of being categorically related. Again, subjects begin at list lengths of 1 and subsequently progress until performance is at or below 75% accuracy. Span is determined the same way as for the category probe task.

Patient Background

Patients were selected on the basis that they were relatively impaired on semantic or phonological STM tasks and based on performance on the aforementioned screening assessments. Namely, patients were only selected if they had intact semantic knowledge and single word processing and were able to complete visual lexical decision, synonymy judgment and pseudohomophone judgment tasks. For measures of semantic and phonological STM (i.e., category and rhyme probe, respectively), we selected patients with spans that were greater than one item and that had variable semantic and phonological STM spans. We ranked the patients by their category probe span for Experiments 1A and 2 and by their rhyme probe span for Experiment 3. Thus, we were able to compare performance on the semantic and phonological versions of the concurrent probe task to their semantic and phonological STM spans. All patients were tested in multiple sessions lasting between 45 minutes to 1.5 hours, typically separated by one week and always separated by at least one day. The patients were tested on Experiments 1A, Experiment 2 and Experiment 3 over the course of 13 sessions (i.e., 5 sessions for experiment 1, 5 sessions for experiment 2, 3 sessions for experiment 3). Each patient will be described in more detail below. A summary of patient's age, years post-stroke, education and location of brain damage is contained in Table 2. Table 3 summarizes patient performance on the various screening assessments and short-term memory tasks.

Patient DW

Patient DW is a 57 year-old female with 16 years of education. In 2000, she suffered a cerebral vascular accident (CVA) primarily involving the LIFG and LMTG. She has relatively intact semantic knowledge (she is 2% less accurate on the PPT than the lower range of controls) as indicated by performance on the PPT and PWM (92 and 95 percent correct, respectively) and performs well on visual lexical decision tasks, as evidenced by her score of 95% accuracy on the visual lexical decision task. She has a category probe span of 2 items and a rhyme probe span of 4 items.

Patient LC

Patient LC is a 66 year-old male with 11 years of education and received his GED. In 2006 he suffered a CVA primarily involving the left parietal lobe. He has intact semantic knowledge as indicated by his performance on the PPT and PWM tasks (100 and 98 percent correct, respectively) and scored 98% correct on the visual lexical decision task. LC has a category probe span of 2.71 items and a rhyme probe span of 2 items.

Patient PP

Patient PP is a 67 year-old male with 20 years of education. In 2011, he suffered a CVA primarily involving the left superior temporal gyrus and insula. He has intact semantic knowledge as indicated by his performance on the PPT and PWM tasks (98 and 99.5 percent correct, respectively). He received a score of 99% accuracy on the visual lexical decision task. He has a category probe span of 3.5 items and a rhyme probe span of 4 items.

Patient RI

Patient RI is an 86 year-old male with 12 years of education. He suffered a CVA in 2005 primarily involving the left temporal lobe and portions of the left parietal lobe. He has intact semantic knowledge as indicated by his performance on the PPT and PWM tasks (98 and 98 percent correct, respectively). He received a score of 94% accuracy on the visual lexical decision task. He has a category probe span of 3 items and a rhyme probe span of 1.67 items.

Experiment 1A-1C

Experiment 1A

Experiment 1A evaluated the relationship between semantic and phonological STM deficits and semantic maintenance. In doing so, Experiment 1A assessed whether the concurrent probe paradigm could be used to assess semantic STM deficits in patients. As mentioned previously, two hypotheses have been put forward concerning the source of deficits in the short-term retention of semantic information. First, patients might have difficulty retaining semantic information due to difficulty inhibiting irrelevant verbal information (Hamilton & Martin, 2005). Alternatively, patients might have difficulty retaining semantic information due to overly rapid decay of that information (Martin & Lesch, 1996). The concurrent probe paradigm has an advantage over other measures of semantic STM (e.g., category probe) in that it can distinguish these two hypotheses regarding the source of semantic STM deficits. Additionally, if the concurrent probe paradigm is useful in measuring semantic STM deficits in patients, then we expect to see a relationship between category probe span and performance on Experiment 1A.

Method: Experiment 1A.

Participants.

Age-matched controls. Eleven neurally healthy older adults were tested on Experiment 1A. One subject was not included in analyses and was removed from the study based on her responses on the post-experiment survey (Appendix D). Specifically, she reported using a phonological maintenance strategy to maintain the maintenance word, despite being instructed to use semantic maintenance to maintain the maintenance word. For the ten age-matched control subjects included in the analyses, mean age was 69.7 years and the average number of years of education was 18.3 years. Although the average education for control subjects was slightly higher than for patients, this did not predict performance on the task (see Results for Experiment 1A). Subjects were recruited from the Brain and Language Lab older controls database and compensated $10 per hour for participation. Informed consent was obtained at the beginning of the experiment and a debriefing was given at the end of the experiment. Debriefing included giving the subjects a brief overview of the purpose of the study in evaluating STM deficits in stroke patients, but no information was given that would compromise the experimental manipulations for the following experiments.

Patients. The four patients outlined above were tested on Experiment 1A. All patients were at least 12 months post-stroke and were classified as aphasic. The average age of patients was 69 years-old and the average number of years of education was 14.75 years. As a reminder, patients were selected based on performance on STM measures, the Pyramids and Palm Trees picture subtest (Howard and Patterson, 1992), single picture-word matching (PWM) and visual lexical decision (for reference, see Tables 1, 2 and 3).

Apparatus. The experiment was administered on an iMac desktop computer running PsyScope (Cohen, MacWhinney, Flatt & Provost, 1993). A button box recorded responses and response times for all events in each trial. Errors in target judgments or probe lexical decision trials were coded using raw data output from Psyscope.

Materials. Materials for Experiment 1 can be found in Appendix A. A total of 90 triplets were used in Experiment 1; 10 were created for this experiment and 80 were taken from Shivde and Anderson (2011). The triplets were created such that each contained a maintenance item (e.g., build), a target item (e.g., erect) and a semantically related critical probe item (e.g., construction). Each testing session was matched for total number of yes and no target item responses and lexical decision item responses. Additionally, each testing session had the same number of semantically related, phonologically related and unrelated probe items. Further, a total of 40 filler trials were used and evenly distributed into each testing session. To create unrelated probe items, semantically related probes were assigned to unrelated maintenance items. All maintenance words were presented a total of 3 times (only once within a session) so that they were paired with each probe type (i.e., semantically related, phonologically related and unrelated). The maintenance and target items were synonyms that were 2-11 letters long (M=5.7), with an average concreteness rating of 353, with 100 being abstract and 700 being concrete (obtained via the MRC Psycholinguistic Database; Wilson, 1988), and an average frequency of 115 (range: 0-1360; Kucera & Francis, 1967). The semantically related probe items were 3-9 letters long (M=5.3), with an average frequency of 120 (range: 0- 1635; Kucera & Francis, 1967). Latent semantic analysis was used to determine the association strength of the probe items to the maintenance word (Landauer, Foltz & Laham, 1998). The average association of the probe word to the maintenance word was .36 (range: .03- .85), where 0 indicates that the words are not associated and 1 indicates that it is the same word.

Procedure. Subjects were first given several practice exercises to familiarize them with the task. First, subjects completed a practice synonymy judgment task. In this practice, subjects were presented with a word in red that stayed on the screen for 2 seconds. Following the red word, a blue word appeared on the screen. Subjects were required to push the yellow button on the button box if the two words meant the same thing. If the two words meant different things, subjects were required to press the green button on the button box. After their decision had been made, “Ready?” appeared on the screen and the experimenter pushed the space bar to continue to the next trial. Subjects completed 10 practice trials of this task and were given verbal feedback on their performance after each trial. Next, subjects completed 10 trials of a practice lexical decision task. In this task, subjects were required to push the yellow button if the presentation was a word, and the green button if the presentation was a non-word. All trials of the lexical decision task were printed in black. Feedback was given at the end of the 10 trials. Subjects then completed a practice version of the concurrent probe task. Subjects were instructed to remember the red maintenance word by focusing on its meaning, not its sound. Then, the red maintenance word appeared and remained on the screen for 2 seconds followed by a reminder (Remember that word!). After the maintenance word was presented, a series of black words appeared on the screen (i.e., the lexical decision filled delay task). Subjects were told to push the yellow button if the presentation was a word, and the green button if it was not. Subjects were told that at some point during the lexical decision task a blue target item would appear. They were instructed to push the yellow button if the target item meant the same thing as the maintenance word and the green button if it did not. In this practice, the experimenter explained the task to the subject step-by-step before the task and reminded them of what to do as they were completing the task. At the end of the last lexical decision trial, "Ready?" appeared on the screen. The experimenter pressed the spacebar at this point to proceed to the next trial. Subjects completed 2 practices of this type. The last practice trials were performed in the exact same way as the actual task. Before beginning this practice, subjects were reminded to focus on the meaning of the red maintenance word in order to retain it during the maintenance interval, not its sound. The experimenter pressed the space bar to begin the task. A red maintenance item appeared on the screen for 2 seconds. The lexical decision task immediately followed presentation of the red maintenance item. Embedded in the lexical decision task was the blue target item. When this item appeared, subjects pressed the yellow button if it meant the same thing as the red maintenance word and the green button if it did not. After each trial, "Ready?" appeared on the screen. The experimenter pushed the space bar when they were ready to move on to the next trial. Although subjects had unlimited time to respond to the blue target item, lexical decision trials would proceed after 2.5 seconds (5 seconds for patients) if a decision had not been made via button press. There were 4 practice trials of this type.

After all practice had been completed, the experimenter gave the subjects 3 reminders. First, subjects were reminded that if they did not make a lexical decision within 2.5 seconds (5 seconds for patients) that the task would automatically progress to the next item and to continue with that item. Second, subjects were told to respond as quickly as possible without making errors. Finally, subjects were reminded to remember the red maintenance word by focusing on its meaning, not its sound. The experimenter then pressed the space bar to begin the experimental trials. In the experimental trials, the red maintenance word would appear on the screen for 2 seconds. The task would then progress to the lexical decision trials. The subjects used the same button presses as in practice and thus the task would advance through the lexical decision trials automatically. Again, if no response was made within 2.5 seconds (5 seconds for patients), then the task would automatically progress to the next trial. When the blue target word appeared, subjects used the same button press as in practice to respond whether it was synonymous with the red maintenance word. At the end of the experimental trial, the word "Ready?" appeared on the screen. Age-matched controls were allowed to press the space bar to proceed to the next trial. For patients, the experimenter remained in the room and pressed the space bar for them when they were ready to move on to the next trial. All age-matched controls and aphasic patients completed a total of 309 trials. Age-matched controls completed these trials in three sessions separated by at least 1 day while patients completed these trials in five sessions separated by at least 1 day.

Post-experiment survey. After the last experimental session, subjects completed a post-experiment survey. This survey asked the subject to describe what strategy they used to maintain the maintenance word, whether they changed their strategy as the target word was further from the maintenance word and any perceptions they had on what the task might be testing. The post-experiment survey is included in Appendix D.

Analyses. For the purposes of interpreting the results of Experiment 1A, we ranked the patients based on their performance on the category probe tasks. Ranked from greatest span to lowest span on the category probe task, patients are ordered as follows: PP, RI, LC and DW. We thus expected that any semantic relatedness effect would be greatest for PP and RI and smallest for LC and DW.

Additionally, we created a variable that represented the difference between semantically related and unrelated probes in the *during* (i.e., semantically related *during* minus unrelated *during*) compared to the *after* (i.e., semantically related *after* minus unrelated *after*) trial type. A positive value indicates an effect whereby the difference between semantically related and unrelated items was larger in the *during* than the *after* trial type. When this occurs, and the difference between semantically related words and unrelated words is significantly different in the *during* but not *after* trial type, such corresponds to the pattern of semantic relatedness effects observed by Shivde and Anderson (2011). While a negative value for this variable still reflects an interaction between trial type and relatedness, it is in the opposite direction as that predicted by Shivde and Anderson (2011) with a greater effect in the *after* condition. Henceforth, this variable will be referred to as the positive or negative semantic relatedness by trial type effect.

Results: Experiment 1A.

Errors. Errors were counted as trials in which the subject inaccurately determined whether the target item was synonymous with the maintenance item. Items were also excluded from data analysis if the subject made an error on the lexical decision for the probe item. Additionally, if the response time for the lexical decision to the probe item was less than 300 ms then that item was excluded. Finally, any response time that was more than 3 standard deviations from each individual's mean response time was excluded from analyses. For controls and patients, Tables 4 and 5 present the probe errors and target errors, respectively, broken down by relatedness and trial type condition (semantically related *during*, unrelated *during,* semantically related *after*, unrelated *after*).

Table 4

Errors on lexical decision trials for probe items by relatedness and trial type, Experiment 1A

Subject	SRD	URD	SRA	URA
Control Average	.6 (.01)	.4 (.01)	.1 (.002)	.5 (.01)
Control Range	0-3 (0-.07)	0-2 (0-.04)	0-1 (0-.02)	0-4 (0-.09)
DW	0 (0)	1 (.02)	1 (.02)	0 (0)
LC	0 (0)	0 (0)	9 (.21)	0 (0)
PP	0 (0)	0 (0)	0 (0)	0 (0)
RI	3 (.07)	5 (.11)	4 (.09)	3 (.07)

Note. The proportion of errors is in parentheses. For *during* trial types, proportion errors are out of 45; for *after* trial types, proportion errors are out of 43. SR=semantically related, UR=unrelated, D=*during*, A=*after.*

Table 5

Errors on synonymy judgment for target items by relatedness and trial type, Experiment 1A

Subject	SRD	URD	SRA	URA
Control Average	3.4 (.08)	3.8 (.08)	2.3 (.05)	3.4 (.08)
Control Range	0-13 (0- .29)	1-9 (.02-.2)	0-10 (0-.23)	1-7 (.02-.16)
DW	8 (.18)	7 (.16)	4 (.09)	8 (.19)
LC	13 (.29)	12 (.27)	6 (.14)	6 (.14)
PP	7 (.16)	7 (.16)	3 (.07)	4 (.09)
RI	8 (.18)	9 (.2)	4 (.09)	8 (.19)

Note. The proportion of errors is in parentheses. For *during* trial types, proportion errors are out of 45; for *after* trial types, proportion errors are out of 43. SR=semantically related, UR=unrelated, D=*during*, A=*after.*

For controls, a total of 180 data points were excluded, accounting for 10% of the total data, and there was no effect of relatedness or trial type for probe errors or target errors. For DW, LC, PP and RI, 19%, 23%, 18% and 28% of their data were removed, respectively. As with the control subjects, DW, PP and RI did not show an effect of relatedness or trial type for probe errors, but LC made more errors to probe items in the semantically related, *after* trial type and was well outside the range of controls (i.e., 0 to 1). Given that he did not make any probe response errors in the *during* trial type, the condition of interest, this result was unexpected and may have been a spurious effect.

With target errors, patients tended to exhibit a larger difference in the number of errors *during* than *after* maintenance, as compared to controls. For controls, the average difference in the number of errors for *during* vs. *after* was 1.2 (i.e., number of errors *during* minus number of errors *after*) and the range was -4, reflecting a larger number of errors in the *after* trial type, to 15. For DW, LC, PP and RI, the difference was 8, 27, 16 and 11, respectively. Patients LC and PP are outside of the range of controls for target errors *during* compared to *after*, although PP is barely outside the range. It is not surprising that the patient's might have slightly more difficulty in the *during* trial type given their relatively impaired semantic STM, but more importantly, there does not appear to be an interaction of relatedness and trial type in their error rates.

Response times: age-matched controls. Only correct response times were included in analyses. A correct response was based on the criteria specified above and was intended to ensure that subjects were accurately maintaining the maintenance item. Table 6 presents the response time data.

Table 6

Average response times (ms) for lexical decision to probe items by relatedness and trial type, Experiment 1A

Relatedness	Semantic, *During*	Unrelated, *During*	Semantic-Unrelated, *During*	Semantic, *After*	Unrelated, *After*	Semantic-Unrelated, *After*
Control Average	880	801	79*+	751	766	-15*
DW	1210	1255	-45	1134	1138	-4
LC	885	918	-33	845	835	10
PP	971	848	123*	766	736	30
RI	1748	1569	179*	1500	1456	44

Note. A * symbol represents that the difference was statistically significant ($p < .05$) in independent samples t-tests, for patients, or paired-samples t-tests by items, for controls. A + symbol represents that the difference was statistically significant ($p < .05$) in paired-samples t-tests by subjects.

Response time data for lexical decision to the probe words (see Table 6) were analyzed using a 2x2 ANOVA with relatedness (semantically related vs. unrelated) as a within-subjects and within-items variable and trial type (*during* vs. *after*) as a within-subjects and between-items variable. The main effect of trial type was significant, $F_1(1, 9) = 7.22$, MSE = 9268, $p = .025$ and $F_2(1, 86) = 61.80$, MSE = 4074, $p < .001$ with longer response times in the *during* trial type than *after* trial type (841 ms vs. 758 ms, respectively). The main effect of relatedness was marginally significant by subjects, $F_1(1, 9) = 4.04$, MSE =9 269, $p = .075$, and highly significant by items, $F_2(1, 86) = 15.33$, MSE = 4074, $p < .001$, reflecting overall longer response times for semantically related probe items than unrelated probe items (816 ms vs. 784 ms, respectively). Critically, the interaction of relatedness and trial type was significant, $F_1(1, 9) = 5.95$, MSE = 3695, $p = .037$ and $F_2(1, 86) = 20.22$, MSE = 4074, $p < .001$. Paired samples t-tests compared semantically related vs. unrelated probe response times for the *during* and *after* trial types. Significantly longer response times for semantically related (M = 880 ms) than unrelated probes (M = 801 ms) were obtained in the *during* trial type, $t_1(9) = 2.27$, p =.05; $t_2(44) = 4.89$, $p < .001$, whereas in the *after* trial type, there was a trend in the opposite direction with shorter times for semantically related (M = 751 ms) than unrelated probes (M = 766 ms), but this difference only reached significance in the analysis by subjects, , $t_1(9) = -2.41$, $p = .039$; $t_2(42) =.60$, $p_2 > .55$. Thus, controls age-matched to the patients demonstrated the pattern of semantic relatedness effects reported by Shivde & Anderson for younger subjects (M = 850 ms vs. 764 ms for semantically related vs. unrelated *during* maintenance, but no significant difference *after, M* = 751 ms vs. 766 ms for semantically related vs. unrelated, respectively; see Figure 2).

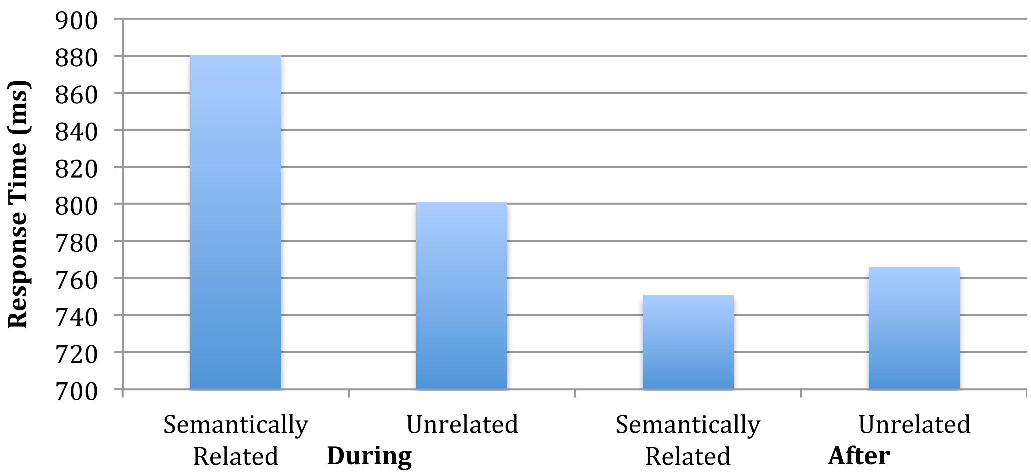

Figure 2. Mean response times for lexical decision to semantically related and unrelated probes for age-matched controls for the *during* and *after* trial types in Experiment 1A.

Response times: patients vs. controls. The control range for the semantic relatedness by trial type interaction effect was -22 ms to 341 ms (M = 94 ms, SD = 129 ms). For patients DW, LC, PP and RI, the value was -41 ms, -43 ms, 91 ms and 135 ms, respectively. Importantly, our patients who were most impaired on the category probe task, DW and LC, performed outside of the range of controls on this measure. In contrast, PP and RI performed within the range of controls. Additionally, patient performance on Experiment 1A, as indexed by their semantic relatedness by trial type interaction effect, was related to their category probe span, but not their rhyme probe span. This can be seen in Figures 3A and 3B. Table 6 contains a summary of patient and control response time data for Experiment 1A.

Further, given the average number of years of education of our controls was quite high, we decided to perform a correlation analysis between the number of years of education and the semantic relatedness by trial type effect in order to rule out the possibility that patient performance differed simply due to education factors. The correlation for controls was not close to significance, $r\ (8) = -.22$, $p > .5$.

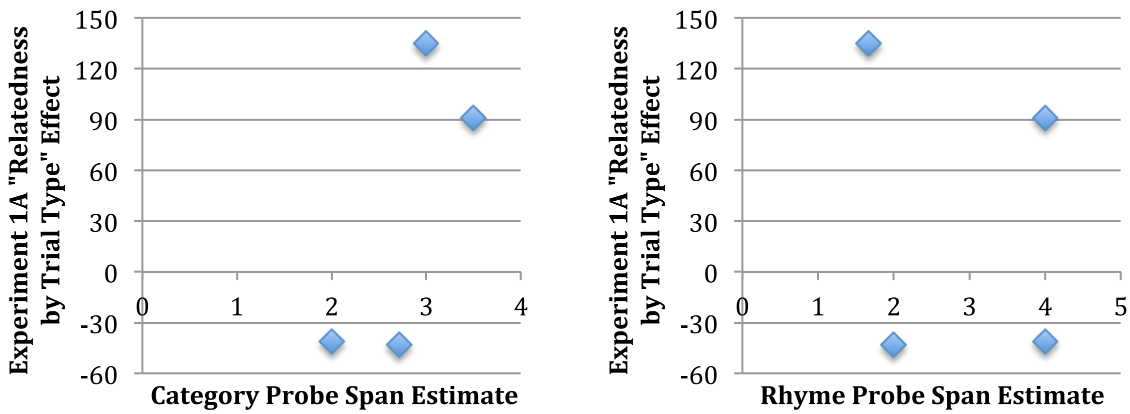

Figure 3. Relationship between probe span and interactions of relatedness by *during* vs. *after* for patients in Experiment 1A. Figure 3A (left) presents the relationship between the category probe span and interactions of relatedness by *during* vs. *after* for patients in Experiment 1A. Figure 3B (right) presents the relationship between the rhyme probe span and interactions of relatedness by *during* vs. *after* for Experiment 1A. As can be seen, performance on Experiment 1A was related to category probe span, but not rhyme probe span.

Response times: patients. Response time data for lexical decisions to the probe words were analyzed individually for each patient (see Table 6). As with controls, response times for error trials were not included in analyses. The patient response time data were initially analyzed on an individual subject basis using a 2x2 ANOVA with relatedness (semantically related vs. unrelated) as a within-items variable and trial type (*during* vs. *after*) as a between-items variable. None of the interaction effects were significant for any of the patients. Presumably, this was due to the large number of errors, which would lead to the elimination of data in both the related and unrelated conditions when an error occurred on either. Thus, independent-samples t-tests were used to analyze the data to increase power in detecting significant effects. Each patient was analyzed individually. DW did not show any significant effect of relatedness in the *during*, p = .66, or *after*, p= .97, trial type and neither did LC, *during* p = .47*, after* p = .77. PP showed a significant effect of relatedness in the *during* trial type, $t(53.04) = 2.22$, $p = .03$, reflecting relatively longer response times for semantically related compared to unrelated word (971 ms vs. 848 ms), but did not show this effect in the *after* trial type, $p = .44$. For *during* trial types, RI took significantly longer to respond to semantically related (1748 ms) compared to unrelated probe items (1569 ms), $t(65) = 2.24$, $p = .029$. The effect of relatedness in the *after* trial type was not significant for RI, $p = .48$.

Discussion: Experiment 1A. The results from the age-matched controls replicated the findings of Shivde and Anderson (2011) with younger adults. Controls exhibited the semantic relatedness effect, with significantly increased response times for semantically related compared to unrelated probes words in the *during*, but not in the *after*, trial type. In contrast, only two patients, RI and PP, exhibited positive interactions of relatedness by *during* vs. *after* trial type that were statistically significant and within the range of controls. Thus, RI and PP both exhibited semantic relatedness effects. DW and LC, who did not exhibit a semantic relatedness effect, had negative semantic relatedness by *during* vs. *after* trial type effects that were below the range of controls. Critically, the size and direction (i.e., positive or negative) of the relatedness by trial type effect was related to the patient's category probe span, where greater category probe span corresponded to semantic relatedness effects and positive interactions of relatedness by *during* vs. *after* trial type that were within the range of controls and smaller category probe spans corresponded to negative interactions of relatedness by *during* vs. *after* trial type that were below the range of controls. Thus, performance on Experiment 1A provides converging evidence for semantic retention deficits with a novel paradigm that has a load of only one item.

Additionally, the results helped disentangle two potential sources of semantic STM deficits: inhibition deficits or overly rapid decay. The data from Experiment 1A supported the hypothesis that DW and LC have difficulty with semantic STM due to overly rapid decay. That is, they cannot maintain items in semantic STM because the item decays more rapidly than in neurally healthy age-matched controls, creating a situation where the maintenance item was not activated to a normal level (i.e., the level to which our control subjects were activating the maintenance item). However, as mentioned previously, we cannot definitively rule out phonological maintenance strategies as an explanation for the lack of a difference in response times for semantically related and unrelated probes in the *during* trial types for DW and LC and it is possible that subjects with semantic STM deficits recruited phonological maintenance strategies to aid in their performance of Experiment 1A given the difficulty they had maintaining items via a semantic strategy (see Experiment 2 for a follow up on this question). Additionally, given the difference in time allowed to make a response on lexical decision trials (i.e., 2.5 seconds for controls and 5 seconds for patients), the amount of time that passed between the maintenance item and the target item was inherently different between patients and controls, and controls might not exhibit a semantic relatedness effect under the longer times that the patients were exposed to. That is, if the difference between controls and patients in the amount of time that had passed between the maintenance item and the probe item was a factor in whether or not a subject exhibited a semantic relatedness effect, then increasing the amount of time that passes between the maintenance item and probe item will eliminate the semantic relatedness effect in controls. Finally, there is a possibility that the semantic relatedness effect was item dependent, meaning that this effect would not transfer to a different set of maintenance items (or, as in Experiment 1C, to *after* items tested in a *during* trial type) and is somehow related only to the items tested in Experiment 1A. Thus, Experiments 1B and 1C, respectively, followed up on these potential problems of Experiment 1A. These experiments are described below, followed by Experiment 2.

Experiment 1B

As mentioned in the procedure of Experiment 1A, the lexical decision task was timed so that if no response was made within a certain amount of time the task would progress automatically. For controls, this time was 2.5 seconds, but for patients, this time was 5 seconds. Thus, if a participant did not respond to any lexical decision trials, the amount of time between the maintenance word and target word in an *after* trial was 5 seconds for controls and 10 seconds for patients; for a *during* trial, the amount of time between the maintenance and target words was 22.5 seconds for controls and 45 seconds for patients. Although there was never a circumstance where no responses were given for any of the lexical decision trials (i.e., a subject passively sat and let the lexical decision trials progress automatically without ever making a response), it is possible that patients may have had more difficulty maintaining the red maintenance item than control subjects simply because they had to maintain it for a longer period of time. To evaluate this possibility, two control subjects were recruited to participate in a "time adjusted" version of the task. In this version of the task, an additional 437 ms was added at the end of each lexical decision trial based on the average amount of time it took patients vs. controls to respond to lexical decision trials. In doing so, we hoped to make the control version of the task more equivalent to the patient version in terms of the amount of time the subject was required to maintain the maintenance word. If the amount of time for maintenance were a relevant factor, we would expect to see a significant difference between control performance on Experiment 1A and Experiment 1B. Accordingly, we would not be able to make any strong conclusions regarding patient data because the task would be deemed inherently different in terms of difficulty. Alternatively, if the amount of time the subject was required to maintain the maintenance word did not effect control performance on the task, we would not expect to see a significant difference between control performance on Experiment 1A and Experiment 1B. This distinction is important to be able to draw conclusions regarding the differences between patient and control performance on Experiment 1A.

Method: Experiment 1B.

Participants. Two age-matched controls were run on Experiment 1B. These controls did not participate in Experiment 1A, but one of the subjects did participate in Experiment 3. For the two subjects, their mean age was 63 years and the average number of years of education was 16. Subjects were recruited from the Brain and Language Lab older controls database and compensated $10 per hour for participation. Informed consent was obtained at the beginning of the experiment and a debriefing was given at the end of the experiment. Debriefing included giving the subjects a brief overview of the purpose of the study in evaluating STM deficits in stroke patients, but no information was given that would compromise the experimental manipulations for the following experiments.

Apparatus. The experiment was run on the same equipment as Experiment 1A.

Materials. The materials were the same as Experiment 1A.

Procedure. The procedure was the same as Experiment 1A with the exception mentioned above. In lieu of automatic progression to the next lexical decision trial after a decision had been made, 437 ms intervening time was placed between each trial. Thus, after each button press, 437 ms passed before the next lexical decision item appeared on the screen.

Post-experiment survey. The post-experiment survey was the same as in Experiment 1A and can be found in Appendix D.

Results: Experiment 1B. Response times were excluded based on the criteria described in Experiment 1A. A total of 53 data points were removed, accounting for 15% of the data (see Tables 7 and 8).

Table 7

Errors on lexical decision trials for probe items by relatedness and trial type, Experiment 1B

Subject	SRD	URD	SRA	URA
1	0 (0)	0 (0)	0 (0)	0 (0)
2	1 (.02)	0 (0)	0 (0)	0 (0)

Note. The proportion of errors is in parentheses. For *during* trial types, proportion errors are out of 45; for *after* trial types, proportion errors are out of 44. SR=semantically related, UR=unrelated, D=*during*, A=*after.*

Table 8

Errors synonymy judgment for target items by relatedness and trial type, Experiment 1B

Subject	SRD	URD	SRA	URA
1	5 (.11)	8 (.18)	14 (.32)	10 (.23)
2	0	7 (.16)	0 (0)	2 (.05)

Note. The proportion of errors is in parentheses. For *during* trial types, proportion errors are out of 45; for *after* trial types, proportion errors are out of 44. SR=semantically related, UR=unrelated, D=*during*, A=*after.*

Response time data for lexical decision to the probe words were analyzed using 2x2 ANOVA with relatedness (semantically related vs. unrelated) as the within-items variable and trial type (*during* vs. *after*) as the between-items variable. A subject analysis was not performed because there were only 2 controls that completed this task and hence power would be too low to detect any effects. The main effect of relatedness was not significant, $p > .2$. The main effect of trial type was significant, $F(1, 81) = 20.78$, MSE = 16015, $p < .001$, reflecting relatively longer response times in the *during* trial type compared to the *after* trial type (721 ms vs 638 ms, respectively). Critically, there was a significant interaction of relatedness and trial type, $F(1, 81) = 4.94$, MSE = 13298, $p = .03$. Paired samples t-tests revealed that the effect of relatedness was not significant in the *after* trial type, $p > .2$ but it was marginally significant in the *during* trial type, $t(41) = 1.91$, $p = .06$. These findings thus replicates the findings of Experiment 1A, though at a marginal level of significance for the relatedness effect in the *during* trial type. These data are presented in Figure 4.

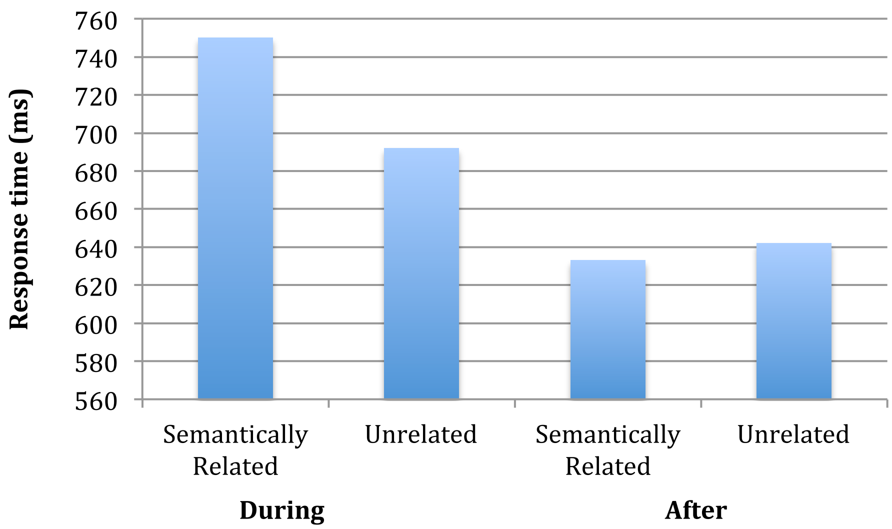

Figure 4. Average response times for lexical decision to semantically related and unrelated probes by age-matched controls for the *during* and *after* trial types in Experiment 1B. The data were averaged across subjects.

Further, we sought to determine whether the results of Experiment 1B were significantly different from those of Experiment 1A. To compare the two experiments, a 2x2x2 mixed ANOVA was run with experiment (1a vs. 1b) and relatedness (semantically related vs. unrelated) as within-items variables and trial type (*during* vs *after*) as between-items variable. Importantly, the three-way interaction of experiment by relatedness by trial type was not close to significance, $p > .9$.

Discussion: Experiment 1B. The results of Experiment 1B indicated that adding the difference in the amount of time the patients took to respond to the lexical decision trials did not significantly affect the way controls performed the tasks. This was indicated by the significant semantic relatedness x trial type interaction, the marginally significant semantic relatedness effect for the *during* trial type, and the non-significant three-way interaction between experiment, relatedness and trial type. These findings support the notion that the conclusions drawn from Experiment 1A regarding the difference in performance for patients compared to controls were not simply due to the amount of time patients had to maintain the maintenance word relative to controls.

Experiment 1C

To determine the extent to which the particular items may have influenced the results of Experiment 1A, a final version of Experiment 1 was run with control subjects. In Experiment 1A, different items were used in the *during* and *after* trial types so that subjects would not see the same items too many times. In Experiment 1C, items that had previously been used for the *during* trial type were switched with those that had previously been used in the *after* trial type. Thus, *during* items were now *after* items, and vice versa. This manipulation sought to ensure that the pattern of results from Experiment 1A would hold when the items were switched. To the extent that the effects seen in Experiment 1A were item independent, we expect to see no significant differences between Experiment 1C and Experiment 1A.

Method: Experiment 1C.

Participants. Three age-matched controls were tested on Experiment 1C. These subjects also participated in Experiment 1A. Mean age was 74.3 years old and average number of years of education was 15.3 years Subjects were recruited from the Brain and Language Lab older controls database and compensated $10 per hour for participation. Debriefing included giving the subjects a brief overview of the purpose of the study in evaluating STM deficits in stroke patients, but no information was given that would compromise the experimental manipulations for the following experiments.

Apparatus. The experiment was run on the same equipment as Experiment 1A.

Materials. The materials were the same as Experiment 1A with the exception noted above. Namely, the maintenance and target items were switched between the *during* and *after* trial types so that items that were in the *during* trial type before were now in the *after* trial type, and vice versa.

Procedure. The procedure was the same as Experiment 1A.

Post-experiment survey. The post-experiment survey was the same as in Experiment 1A and can be found in Appendix D.

Results: Experiment 1C. Response times were excluded based on the criteria described in Experiment 1A. A total of 45 data points were removed, accounting for 8.5% of the data.

Response time data for lexical decision to the probe words were analyzed using 2x2 ANOVA with relatedness (semantically related vs. unrelated) as the within-items variable and trial type (*during* vs. *after*) as the between-items variable. A subject analysis was not performed because there were only 3 controls that completed this task and hence power would be too low to detect any effects. The main effect of relatedness was marginally significant, $F(1, 86) = 3.61$, MSE = 16153, $p = .061$, reflecting relatively longer response times for semantically related compared to unrelated items (854 ms vs 817 ms, respectively). The main effect of trial type was significant, $F(1, 86) = 37.53$, MSE = 17611, $p < .001$, reflecting relatively longer response times in the *during* trial type compared to the *after* trial type (897 ms vs 774 ms, respectively). Even though the relatedness effect was 49 ms in the *during* condition and 24 ms in the *after* condition the interaction was not significant, $p > .5$. These data are presented in Figure 5.

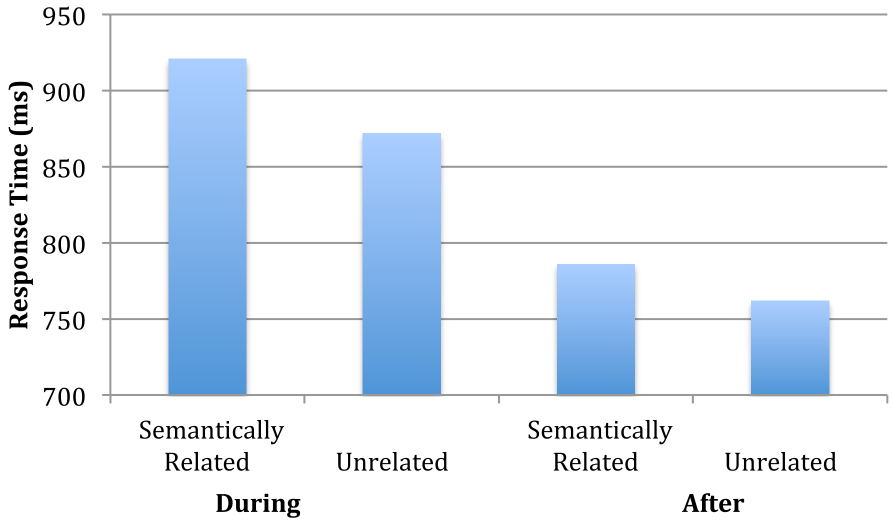

Figure 5. Average response times for lexical decision to semantically related and unrelated probes by age-matched controls for the *during* and *after* trial types in Experiment 1C. The data were averaged across subjects.

The lack of a significant interaction between relatedness and trial type was not expected based on the results of Experiment 1A and could be attributed to a lack of power in detecting effects given that Experiment 1A had 10 subjects while Experiment 1C only had 3 subjects. Thus, to determine whether the results of Experiment 1C were significantly different from Experiment 1A, we ran a 2x2x2 mixed ANOVA with relatedness (semantically related vs. unrelated) as a within-items variables and experiment (1a vs. 1c) and trial type (*during* vs *after*) as between-items variables. Importantly, the three-way interaction of experiment by relatedness by trial type was not significant, $p > .17$.

Discussion: Experiment 1C. The lack of a significant interaction between relatedness and trial type in Experiment 1C was surprising, but there are at least two possible explanations. First, there were only 3 subjects in Experiment 1C while in Experiment 1A there were 10 subjects. It is possible that with greater power, the interaction for Experiment 1C would reach significance. Second, all subjects who participated in Experiment 1C had also participated in Experiment 1A. Thus, it is possible that there was some sort of effect of seeing the items so many times that attenuated the relatedness effect. We think this is unlikely given that several weeks had passed between participation in the two experiments, but it may still be a possibility.

Despite the lack of a significant interaction between relatedness and trial type in Experiment 1C, there was no significant difference between the results of Experiment 1C and those of Experiment 1A. In trying to determine whether or not it can be stated that the effect in Experiment 1A can be generalized to other items, we ran a 2 x 2 x 2 mixed ANOVA with experiment (1A vs. 1C) and trial type (*during* vs. after) as between-items variables and relatedness (semantically related vs. unrelated) as within-items variables to see if the semantic relatedness effect would persist; then, we compared the average effect of relatedness in the *during* trial type to the findings of Shivde and Anderson (2011). The interaction of relatedness and trial type was significant, $F(1, 172) = 6.37$, MSE = 63151, $p = .013$, ($M = 895$ ms vs. 832 ms for semantically related vs. unrelated *during* maintenance compared to 769 ms vs. 760 ms, *after*). The finding of a 63 ms difference in the *during* trial type is comparable to, albeit slightly smaller than, the pattern of semantic relatedness effects reported by Shivde & Anderson ($M =$ 850 ms vs. 764 ms for semantically related vs. unrelated *during* maintenance, a difference of 86 ms, but no significant difference *after*) for younger subjects. These results suggest that the semantic relatedness effect found in Experiment 1A was item independent. This was important to establish given the conclusions drawn from Experiment 1A are only valid to the extent that they do not depend on the particular items chosen for the task.

Discussion: Experiment 1A-1C

The results of Experiments 1A-1C support the initial conclusions drawn in Experiment 1A. We replicated the findings of Shivde and Anderson (2011) with older adults and patients RI and PP. We also determined that the finding of a semantic relatedness effect was item independent and that the performance of control subjects did not differ from the performance of patients DW and LC simply due to the amount of time required to maintain the maintenance item.

As mentioned in the discussion of Experiment 1A, it is possible that a lack of the semantic relatedness effect for patients DW and LC resulted from their adopting a phonological maintenance strategy. Such a strategy would be adaptive if they have a semantic STM deficit, since phonological maintenance would allow them to perform the task. However, it is possible that they adopted a phonological maintenance strategy even though they could have used a semantic strategy. Although this would go against the task instructions, even some controls (n = 1, excluded from analyses) adopted a phonological strategy, according to self-report, even though it was emphasized that they were to use semantic maintenance. Thus, in Experiment 2 we wished to determine if there was evidence that these patients used a phonological maintenance strategy.

Experiment 2

Experiment 2 sought to determine whether subjects were using a phonological maintenance strategy despite being instructed to use a semantic maintenance strategy. Thus, in this experiment, the probe items of critical interest were phonologically related to the maintenance word. Given that subjects were instructed to use a semantic maintenance strategy and given their performance on Experiment 1A, we did not expect control subjects or patients PP and RI to show any difference between phonologically related probe items and unrelated probe items in either the *during* or *after* trial types. However, we anticipated the possibility that patients DW and LC may have been using phonological maintenance strategies in Experiment 1A. Shivde and Anderson (2011) found that in a phonological maintenance condition (that is, when subjects had to judge the relation between the maintenance item and the target on phonological grounds), subjects took longer to respond to phonologically related probe items in the lexical decision task compared to unrelated probe items in the *during*, but not the *after*, trial type (i.e., phonological relatedness effect). Thus, any significant phonological relatedness effect in Experiment 2 would indicate that the subject was using a phonological maintenance strategy to maintain the maintenance word. If DW and LC do not show a phonological relatedness effect in this semantic maintenance task, this would argue against the possibility that the absence of a semantic relatedness effect in Experiment 1a was due to their decision to adopt a phonological maintenance strategy, even though they might have adopted a semantic maintenance strategy. If they do show a phonological relatedness effect, then the findings from Experiment 1a would be more ambiguous.

Method: Experiment 2

Participants.

Age-matched controls. Ten neurally healthy older adults were tested on Experiment 2. These were the same control subjects as tested in Experiment 1A. Subjects were compensated $10 per hour for participation. Debriefing included giving the subjects a brief overview of the purpose of the study in evaluating STM deficits in stroke patients, but no information was given that would compromise the experimental manipulations for the following experiments.

Patients. The same four patients were tested on Experiment 2. As a reminder, task and patient summary can be found in Tables 1, 2 and 3.

Apparatus. Experiment 2 was run on the same equipment as Experiment 1.

Materials. Materials for Experiment 2 can be found in Appendix B. A total of 90 triplets were created for Experiment 2. The triplets were created such that each contained a maintenance item (e.g., decide), a target item (e.g., determine) and a phonologically related (rhyming) critical probe item (e.g., ride). Each testing session was matched for total number of yes and no target item responses and lexical decision item responses. Additionally, each testing session had the same number of semantically related, phonologically related and unrelated probe items. Forty filler trials were evenly distributed into each testing session as well. To create unrelated probe items, phonologically related probe items were assigned to unrelated maintenance items. All maintenance words were presented a total of 3 times (only once within a session) so that they were paired with each probe type (i.e., semantically related, phonologically related and unrelated). The maintenance and target items were synonyms that were 3-10 letters long (M=5.8), with an average concreteness rating of 391, where 100 is abstract and 700 is concrete (obtained via the MRC Psycholinguistic Database; Wilson, 1988), and an average frequency of 62 (range: 0-589; Kucera & Francis, 1967). The phonologically related probe items were 3-9 letters long (M=4.7), with an average frequency of 66 (range: 0-676; Kucera & Francis, 1967).

Procedure. Subjects were first given several practice exercises to familiarize them with the task. Practice exercises were the same as Experiment 1.

After all practice had been completed, the experimenter gave the subjects the same 3 reminders as in Experiment 1. All age-matched controls and aphasic patients completed a total of

310 trials. Age-matched controls completed these trials in three sessions separated by at least 1 day while aphasia patients completed these trials in five sessions separated by at least 1 day.

Post-experiment survey. The post-experiment survey was the same as in Experiment 1A and can be found in Appendix D.

Analyses. We predicted that if subjects were using a phonological maintenance strategy, that they would show a phonological relatedness effect. In contrast, if subjects were using a semantic maintenance strategy, as instructed, then they should not show a phonological relatedness effect. Of particular interest were the results for DW and LC, the patients with the smallest category probe spans, who failed to show the semantic relatedness effect in Experiment 1A. We wished to determine whether the failure to show this effect resulted from the adoption of a phonological maintenance strategy.

We created an interaction contrast variable that represented the difference between phonologically related and unrelated probes in the *during* (i.e., phonologically related *during* minus unrelated *during*) compared to the *after* (i.e., phonologically related *after* minus unrelated *after*) trial type. A positive value indicates an effect whereby the difference between phonologically related and unrelated items was larger in the *during* than the *after* trial type. When this occurs, and the difference between phonologically related words and unrelated words is significantly different in the *during* but not *after* trial type, it corresponds to the phonological relatedness effect reported by Shivde and Anderson (2011). A negative value indicated that the difference between phonologically related and unrelated items was smaller in the *during* than the *after* trial type. This variable will be referred to as the phonological relatedness by trial type effect.

Results: Experiment 2

Errors. For controls and patients, Tables 9 and 10 present the probe errors and target errors, respectively, broken down by relatedness and trial type condition (phonologically related *during*, unrelated *during,* phonologically related *after*, unrelated *after*).

Table 9

Errors on lexical decision trials for probe items by relatedness and trial type, Experiment 2

Subject	PRD	URD	PRA	URA
Control Average	.55 (.01)	.45 (.01)	1.2 (.03)	.8 (.02)
Control Range	0-2 (0-.04)	0-2 (0-.04)	0-5 (0-.12)	0-4 (0-.09)
DW	3 (.07)	2 (.04)	2 (.05)	4 (.09)
LC	0 (0)	0 (0)	0 (0)	0 (0)
PP	1 (.02)	1 (0.02)	1 (.02)	2 (.05)
RI	1 (.02)	5 (.11)	1 (.02)	1 (.02)

Note. The proportion of errors is in parentheses. For *during* trial types, proportion errors are out of 46; for *after* trial types, proportion errors are out of 43. PR=phonologically related, UR=unrelated, D=*during*, A=*after.*

Table 10

Errors on synonymy judgment for target items by relatedness and trial type, Experiment 2

Subject	PRD	URD	PRA	URA
Control Average	2 (.04)	2.8 (.06)	1.2 (.03)	2 (.05)
Control Range	0-4 (0-.09)	0-9 (0-.2)	0-4 (0-.09)	0-5 (0-.12)
DW	8 (.17)	12 (.26)	3 (.07)	5 (.12)
LC	14 (.3)	11 (.24)	2 (.05)	6 (.14)
PP	4 (.09)	8 (.17)	2 (.05)	1 (.02)
RI	8 (.17)	8 (.17)	3 (.07)	3 (.07)

Note. The proportion of errors is in parentheses. For *during* trial types, proportion errors are out of 46; for *after* trial types, proportion errors are out of 43. PR=phonologically related, UR=unrelated, D=*during*, A=*after.*

For controls, a total of 148 data points were excluded using the same criteria as Experiment 1A, accounting for 8% of the total data. Neither probe errors nor target errors showed an effect of relatedness or trial type. For DW, LC, PP and RI, 23%, 21%, 14% and 17% of their data were removed, respectively. As with controls, the patient's probe errors showed no effect of relatedness or trial type. For target errors, there was a trend towards slightly more errors in the *during* condition. For controls, the average difference in the number of errors for *during* vs. *after* was .9 (i.e., number of errors *during* minus number of errors *after*) and the range was -6, reflecting a larger number of errors in the *after* trial type, to 13. For DW, LC, PP and RI, the difference was 8, 25, 10 and 9, respectively. Patient LC was outside of the range of controls for target errors *during* compared to *after*. As in Experiment 1A, it is not surprising that the patient's might have slightly more difficulty in the *during* trial type given their relatively impaired semantic STM, but more importantly, there does not appear to be an interaction of relatedness and trial type in their error rates.

Response times: age-matched controls. Response times were included based on the criteria described above. Response time data for lexical decision to the probe words were analyzed using a 2x2 ANOVA with relatedness (phonologically related vs. unrelated) as the within-subjects and within-items variable and trial type (*during* vs. *after*) as the within-subjects and between-items variable. The main effect of trial type was significant both in the subjects analysis, $F_1(1, 9) = 8.85$, MSE = 1510, $p = .016$, and in the items analysis, $F_2(1, 87) = 6.93$, MSE = 9487, $p = .01$, reflecting overall longer response times in the *during* trial type than *after* trial type (800 ms vs. 764 ms, respectively). The main effect of relatedness was not significant by subjects, $p_1 = .17$, but was marginally significant by items, $F_2(1, 87) = 3.24$, MSE = 4567, $p = .075$, reflecting relatively longer response times for phonologically related probe items compared to unrelated probe items (807 ms vs. 788 ms, respectively). Critically, although the phonological relatedness was slightly greater in the *during* condition (16 ms) than in the *after* condition (5 ms) the interaction of relatedness and trial type was not significant by subjects, $p_1 = .32$, or by items, $p_2 = .16$. Figure 6 presents this data in graphical format.

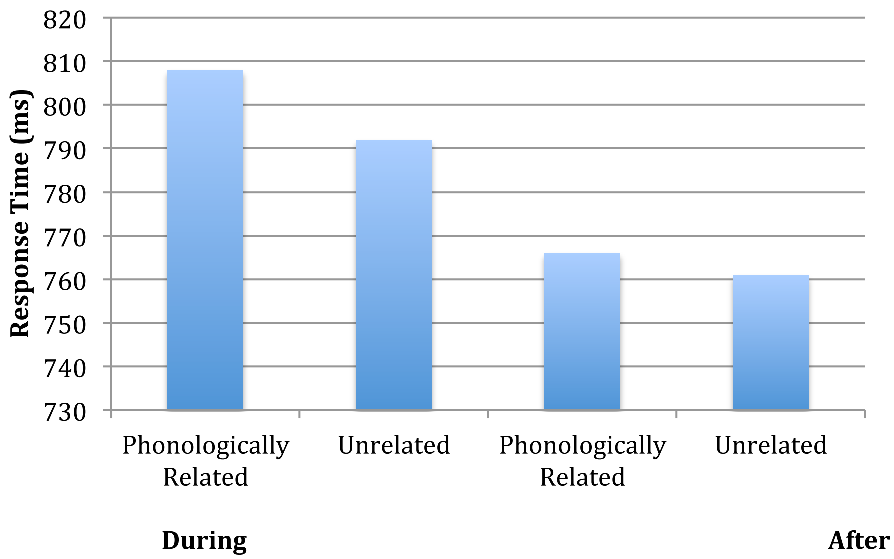

Figure 6. Average response times for lexical decision to phonologically related and unrelated probes by age-matched controls for the *during* and *after* trial types in Experiment 2. The data were averaged across subjects.

Response times: patients vs. controls. The control range for the phonological relatedness by trial type effect was -51 ms to 63 ms (*M*=11, *SD*=34). For patients DW, LC, PP and RI, the value was -71 ms, -104 ms, -47 ms and -247 ms, respectively. Thus, all patients except PP fell outside of the range of controls in a direction that was not predicted (i.e., they all showed negative relatedness by trial type effects when phonological maintenance strategies would predict a positive relatedness by trial type effect). That is, all of the patients showed a smaller phonological relatedness effect in the during condition than in the after condition – the opposite of the pattern that would be predicted for a phonological maintenance strategy based on the results of Shivde and Anderson (2010). Although the patients were consistent in showing a smaller effect in the during condition, they differed considerably in whether the effects in the during or after condition were, on their own, facilitatory or inhibitory. Additionally, there was no clear relation between performance on Experiment 2 and category probe span, but instead there appeared to be a relation to rhyme probe span (see Figures 7A and 7B). Table 11 contains a summary of the data for Experiment 2. Analyses of the effects for individual patients are reported below.

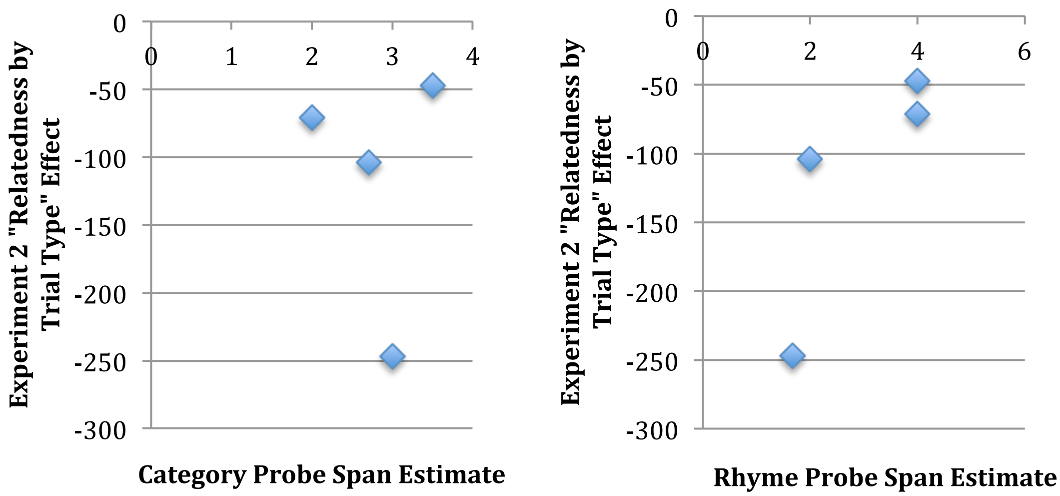

Figure 7. Relationship between probe span and interactions of relatedness by *during* vs. *after* for patients in Experiment 2. Figure 7A (left) presents the relationship between the category probe span and interactions of relatedness by *during* vs. *after* for patients in Experiment 2. Figure 7B (right) presents the relationship between the rhyme probe span and interactions of relatedness by *during* vs. *after* for Experiment 2. As can be seen, performance on Experiment 2 was related to rhyme probe span, but not category probe span.

Table 11

Average response times (ms) for lexical decision to probe items by relatedness and trial type, Experiment 2

Relatedness	Phonological, *During*	Unrelated, *During*	Phonological -Unrelated, *During*	Phonological, *After*	Unrelated, *After*	Phonological -Unrelated, *After*
Control Average	808	792	16	766	761	5
DW	1456	1333	123	1177	982	195*
LC	997	1024	-27	935	857	78*
PP	850	970	-120*	733	806	-73
RI	1843	1973	-130	1856	1739	117

Note. Phonological-unrelated refers to the difference between the average response time for phonologically related and unrelated probe items. A * symbol represents that the difference was statistically significant ($p < .05$) in independent samples t-tests.

Response times: patients. Response time data were included based on the same criteria as Experiment 1A. The patient data were analyzed the same way as Experiment 1A, except where noted. Similarly to Experiment 1A, the patient response time data were initially analyzed using a 2x2 ANOVA with relatedness (phonologically related vs. unrelated) as a within-items variable and trial type (*during* vs. *after*) as a between-items variable. RI was the only patient to exhibit a significant interaction effect, $F(1, 19) = 6.12$, MSE=107422, $p = .023$. He showed evidence of facilitation in the *during* condition, with faster response times for phonologically related (1843 ms) compared to unrelated (1973 ms), but interference in the *after* condition, with slower response times for phonologically related (1856 ms) compared to unrelated (1739 ms), although neither effect was close to significance in independent samples t-test- *during*, $p = .34$, or *after*, $p = .54$. Presumably, there was not enough power in the ANOVA to detect the interaction effects for the other patients due to the large number of errors, so independent-samples t-tests were used to analyze the data to increase power in detecting significant effects. Unexpectedly, PP showed substantially faster times in the phonologically related (850 ms) than unrelated condition (970 ms) in the *during* condition, a difference that reached significance, $t(55.88) = -1.967$, $p = .05$. In the *after* trial type, he also showed faster times in the phonologically related (733 ms) than unrelated condition (806 ms), but this difference was not significant, $p = .11$. For DW, the effect of relatedness was not significant in the *during* trial type, $p = .4$, but was significant in the *after* trial type, $t(57.48) = 2.53$, $p = .01$, reflecting relatively longer response times to phonologically related compared to unrelated items (1177 ms vs. 982 ms, respectively). LC did not show a significant effect of relatedness in the *during* trial type, $p = .68$, but did show a significant effect of relatedness in the *after* trial type, $t(76) = 2.24$, $p = .03$, reflecting relatively longer response times to phonologically related compared to unrelated items (935 ms vs. 857 ms, respectively).

Discussion: Experiment 2

Control subjects did not exhibit a phonological relatedness effect in Experiment 2 as evidenced by a lack of a significant interaction between trial type and relatedness in the predicted direction (i.e., significantly longer response times for phonologically related compared to unrelated items in the *during*, but not *after*, trial type). Critically, this shows that they were not using a phonological maintenance strategy in this task in which they were instructed to hold onto the meaning of the maintenance item. This lack of a phonological relatedness effect contrasts with the significant semantic relatedness effect in Experiment 1A and, together, the results from the two experiments indicate that these subjects were employing semantic maintenance.

None of the patients exhibited a phonological relatedness effect of greater phonological interference in the *during* than the *after* condition – if anything, the patients showed the opposite pattern of a greater phonological interference in the *after* than the *during* condition. Of particular interest were the results from DW and LC as they failed to show the semantic relatedness effect in Experiment 1A and we had hypothesized that they might have been using a phonological maintenance strategy. Unexpectedly, both patients showed a significant phonological relatedness effect in the *after* condition. DW showed a non-significant phonological interference effect in the *during* condition whereas LC showed non-significant facilitation. In relation to our hypotheses about phonological maintenance (i.e., inhibition deficit would predict priming vs. overly rapid decay would predict no difference), the effect shown by these patients did not reflect phonological maintenance strategies. DW and LC took significantly more time to respond to phonologically related compared to unrelated words in the *after* trial type. Further complicating interpretation of the results, PP showed a priming effect in the *during* trial type. Given the lack of a clear pattern of results in relation to our hypotheses, we looked at the interactions of relatedness by *during* vs. *after*.

In looking at the phonological relatedness by trial type effect, all patients exhibited a negative effect (i.e., the difference between phonologically related and unrelated items was smaller in the *during* compared to the *after* trial type). This effect was related to their rhyme probe span (see Figure 7B); if a patient had a larger rhyme probe span, then their phonological relatedness by trial type effect was closer to the range of controls. It is not clear why this would be the case, but it is possible that the difference between the patients' performance and control performance is due to some other factor (e.g., combination of their semantic and phonological STM deficits). This will be discussed further in the General Discussion.

Experiment 3

Experiment 3 evaluated the effects of phonological maintenance on lexical decision to phonologically related probes. We wished to determine if older controls would show a pattern of phonological interference effects like that reported by Shivde and Anderson for younger subjects, with greater phonological interference in the *during* than *after* condition. If the effect could not be replicated in older controls, we would not expect patients to show this pattern. Assuming that such effects are obtained for controls, we wished to determine if patients with greater phonological retention capacities would show the same pattern as controls, whereas those with reduced phonological retention capacities would not show phonological interference effects. It was hoped that in a task with explicit instructions for phonological maintenance, the results would be clearer than those in Experiment 2 and potentially help to clarify those results.

Paired with Experiment 1A, Experiment 3 sought to establish a dissociation between semantic and phonological maintenance effects in the concurrent probe paradigm. While Experiment 1 required participants to make a synonymy judgment between the maintenance word and the target word, Experiment 3 required participants to determine whether the maintenance word would be pronounced exactly the same way as a non-word target item, or, in other words, participants had to determine if a non-word target item was a pseudohomophone of the maintenance item and subjects were instructed to use a phonological maintenance strategy, The probe items of critical interest were phonologically related to the maintenance item in lieu of being semantically related.

Further, if the age-matched control subjects exhibit a phonological relatedness effect in Experiment 3, the dissociation between the effects of Experiment 1A and Experiment 3 (i.e., semantic relatedness effects and phonological relatedness effects, respectively) would provide more support for the notion of separable semantic and phonological maintenance abilities.

Method: Experiment 3

Participants.

Age-matched controls. Ten neurally healthy older adults were tested. Of these subjects, 8 had participated in Experiment 1A and Experiment 2, 1 had participated in Experiment 1B and 1 had been tested on a pilot version of Experiment 1B. For these subjects, the mean age was 69.2 years-old and the average number of years of education was 16.7 years. Subjects were compensated $10 per hour for participation. Debriefing included giving the subjects a brief overview of the purpose of the study in evaluating STM deficits in stroke patients, but no information was given that would compromise the experimental manipulations for the following experiments.

Patients. The same four patients were tested on Experiment 2. As a reminder, a description of each patient and a summary of all tasks can be found in Tables 1, 2 and 3.

Apparatus. The experiment was administered on the same equipment as Experiment 1.

Materials. Materials for Experiment 3 can be found in Appendix C. A total of 90 triplets were created for Experiment 3. The triplets were created such that each contained a maintenance item (e.g., chief), a target non-word item (e.g., cheef) and a phonologically related (rhyming) critical probe item (e.g., beef). Each testing session was matched for total number of yes and no target item responses and lexical decision item responses. Additionally, each testing session had the same number of semantically related, phonologically related and unrelated probe items. Several filler trials were inserted into each testing session as well. To create unrelated probe items, phonologically related probe items were assigned to unrelated maintenance items. All maintenance words were presented a total of 3 times (only once within a session) so that they were paired with each probe type (i.e., semantically related, phonologically related and unrelated). The maintenance item was a word and the target item was a non-word that was a pseudohomophone of the maintenance item (or not). Non-pseudohomophone trials were created the same way as Experiments 1 and 2, namely that targets that were pseudohomophones in one trial were used as non-pseudohomophone targets in other trials. The maintenance items were 3-7 letters long (M=4.5), with an average concreteness rating of 450 where 100 is abstract and 700 is concrete (obtained via the MRC Psycholinguistic Database; Wilson, 1988), and an average frequency of 159 (range: 0-2439; Kucera & Francis, 1967). The non-word target items were 2-9 letters long (M=4.5). The phonologically related probe items were 2-9 letters long (M=4.7), with an average frequency of 134 (range: 0-1895; Kucera & Francis, 1967).

Procedure. Subjects were first given several practice exercises to familiarize them with the task. First, subjects completed a practice pseudohomophone judgment task. In this task, subjects were first presented with a word in red that stayed on the screen for 2 seconds. Following the red word, a blue non-word appeared on the screen. Subjects were required to push the yellow button on the button box if the two items would be pronounced exactly the same way. If the two items would not be pronounced exactly the same way, subjects were required to press the green button on the button box. After their decision had been made, “Ready?” appeared on the screen and the experimenter pushed the space bar to continue to the next trial. Subjects completed 10 practice trials of this task and were given verbal feedback on their performance. Next, subjects completed a practice lexical decision task. In this task, subjects were required to push the yellow button if the presentation was a word, and the green button if the presentation was a non-word. All trials of the lexical decision task were printed in black. After these two practices, subjects completed a practice version of the concurrent probe task. In this practice, the red maintenance word remained on the screen for 2 seconds followed by a reminder (Remember that word!). Subjects were instructed to remember the word by focusing on its sound, not its meaning. After the maintenance word was presented, a series of black words appeared on the screen (i.e., the lexical decision filled delay task). Subjects were told to push the yellow button if the presentation was a word, and the green button if it was not. Subjects were told that at some point during the lexical decision task a blue target item would appear. Subjects were told to push the yellow button if the target item was pronounced exactly the same way as the maintenance word and the green button if it was not. In this practice, the experimenter explained the task to the subject step-by-step before the task and reminded them of what to do as they were completing the task. At the end of the last lexical decision trial, “Ready?” appeared on the screen. The experimenter pressed the spacebar at this point to proceed to the next trial. Subjects completed 2 practices of this type. The last practice trials were exactly the same as the actual task. Before beginning this practice, subjects were reminded to focus on the sound of the red maintenance word in order to retain it during the maintenance interval, not its meaning. The experimenter pressed the space bar to begin the task. A red maintenance item appeared on the screen for 2 seconds. The lexical decision task immediately followed presentation of the red maintenance item. Embedded in the lexical decision task was the blue target item. When this item appeared, subjects pressed the yellow button if it would be pronounced exactly the same way as red maintenance word and the green button if it was not. After each trial, “Ready?” appeared on the screen. The experimenter pushed the space bar when they were ready to move on to the next trial. Although subjects had unlimited time to respond to the blue target item, lexical decision trials would proceed after 2.5 seconds (5 seconds for patients) if a decision had not been made via button press. There were 4 practice trials of this type.

After all practice had been completed, the experimenter gave the subjects 3 reminders. First, subjects were reminded that if they did not make a lexical decision within 2.5 seconds (5 seconds for patients) that the task would automatically progress to the next item and to continue with that item. Second, subjects were told to respond as quickly as possible without making errors. Finally, subjects were reminded to remember the red maintenance word by focusing on its sound, not it’s meaning. All age-matched controls and aphasic patients completed a total of 309 trials. Age-matched controls completed these trials in three sessions separated by at least 1 day. In Experiment 3, aphasic patients also completed these trials in three sessions separated by at least 1 day. We changed the number of sessions the patients had to complete to shorten the number of times they had to come into the lab for testing. In Experiment 1A and Experiment 2, patients were able to complete the experimental session in less than 45 minutes, and with the addition of the extra trials they were able to complete the experimental session in about 1 hour.

Post-experiment survey. The post-experiment survey was the same as in Experiment 1A and can be found in Appendix D.

Analyses. For the purposes of interpreting the results of Experiment 3, we used the same analysis as in Experiment 2.

Results: Experiment 3

Errors. For controls and patients, Tables 12 and 13 present the probe errors and target errors, respectively, broken down by relatedness and trial type condition (semantically related *during*, unrelated *during,* semantically related *after*, unrelated *after*).

Table 12

Errors on lexical decision trials for probe items by relatedness and trial type, Experiment 3

Subject	PRD	URD	PRA	URA
Control Average	.7 (.02)	1 (.02)	.9 (.02)	.6 (.01)
Control Range	0-1 (0-.02)	0-6 (0-.13)	0-4 (0-.09)	0-2 (0-.04)
DW	5 (.11)	5 (.11)	4 (.09)	6 (.13)
LC	0 (0)	0 (0)	0 (0)	0 (0)
PP	2 (.04)	2 (.04)	1 (.02)	1 (.02)
RI	1 (.02)	2 (.04)	1 (.02)	1 (.02)

Note. The proportion of errors is in parentheses. For *during* and *after* trial types, proportion errors are out of 45. PR=phonologically related, UR=unrelated, D=*during*, A=*after.*

Table 13

Errors on pseudohomophone judgment for target item by relatedness and trial type, Experiment 3

Subject	PRD	URD	PRA	URA
Control Average	.7 (.02)	1 (.02)	.8 (.02)	.7 (.02)
Control Range	0-5 (0-.11)	0-4 (0-.09)	0-5 (0-.11)	0-3 (0-.07)
DW	7 (.16)	5 (.11)	7 (.16)	2 (.04)
LC	15 (.33)	15 (.33)	3 (.07)	4 (.09)
PP	5 (.11)	3 (.07)	2 (.04)	1 (.02)
RI	2 (.04)	4 (.09)	2 (.04)	2 (.04)

Note. The proportion of errors is in parentheses. For *during* and *after* trial types, proportion errors are out of 45. PR=phonologically related, UR=unrelated, D=*during*, A=*after.*

For controls, a total of 100 data points were excluded, accounting for 5.6% of the total data. There was no effect of relatedness or trial type for probe errors or target errors. For DW, LC, PP and RI, 24%, 22%, 13% and 9% of their data were removed, respectively. As with controls, there was no effect of relatedness or trial type for probe errors and no effect of relatedness for target errors. For target errors, there was a trend towards slightly more errors in the *during* condition. For controls, the average difference in the number of errors for *during* vs. *after* was .6 (i.e., number of errors *during* minus number of errors *after*) and the range was -2, reflecting a larger number of errors in the *after* trial type, to 2. For DW, LC, PP and RI, the difference was 4, 34, 7 and 2, respectively. Thus, all patients except RI were outside of the range of controls for target errors *during* compared to *after*. As in Experiments 1A and 2, it is not surprising that the patient's might have slightly more difficulty in the *during* trial type given their relatively impaired STM, although it is not clear why RI, the patient with the lowest rhyme probe span, performs within the range of controls while the other patients do not. However, more importantly, there does not appear to be an interaction of relatedness and trial type in their error rates.

Response times: age-matched controls. Response times were included based on the same criteria as Experiment 1A. Response time data for lexical decision to the probe words were analyzed using a 2x2 ANOVA with relatedness (phonologically related vs. unrelated) as the within-subjects and within-items variable and trial type (*during* vs. *after*) as the within-subjects and between-items variable. The main effect of trial type was significant both in the subjects analysis, $F_1(1, 9) = 14.26$, MSE = 1996, $p = .004$, and in the items analysis, $F_2(1, 88) = 12.42$, MSE = 17773, $p = .001$; the significant main effect by subject and by items reflects overall longer response times in the *during* trial type than *after* trial type (839 ms vs. 769 ms, respectively. The main effect of relatedness was not significant by subjects, $p > .1$, but was significant by items, $F_2(1, 88) = 11.38$, MSE = 4015, $p = .001$; this main effect reflects overall longer response times for phonologically related probe items than unrelated probe items (816 ms vs. 791 ms, respectively). Critically, the interaction of relatedness and trial type was significant by subjects, $F_1(1, 9) = 5.66$, MSE = 1996, $p = .041$, and by items, $F_2(1, 88) = 9.67$, MSE = 4015, $p = .003$. In the *during* trial type, the difference between phonologically related and unrelated probe response times was marginally significant by subject and highly significant by items in the *during* trial type, $t_1(9) = 2.17$, $p = .059$; $t_2(44) = 4.06$, $p < .001$, but in the *after* trial type, was not significant by subject or by item, , $p_1 > .3$; $p_2 > .8$ (see Figure 8).

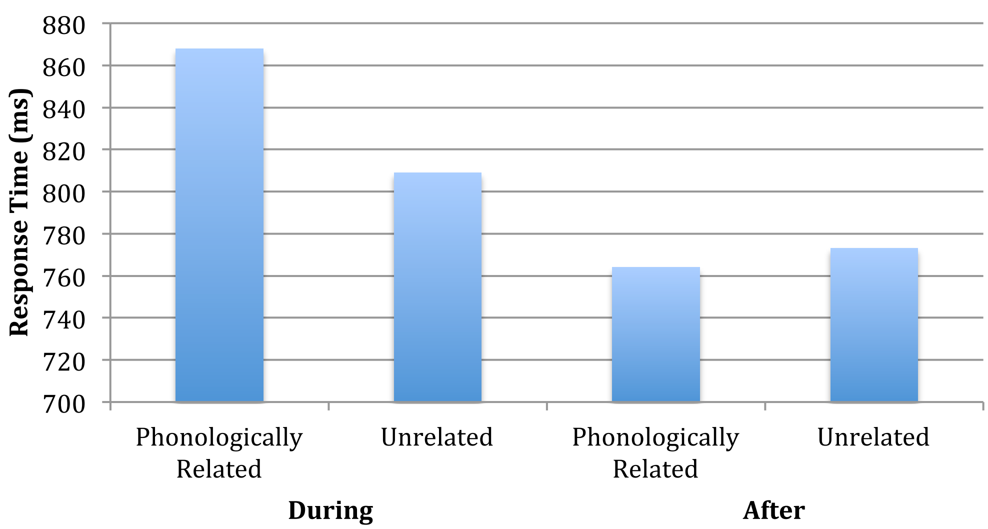

Figure 8. Average response times for lexical decision to phonologically related and unrelated probes by age-matched controls for the *during* and *after* trial types in Experiment 3. The data were averaged across subjects Experiment 3 was related to category probe span, but not rhyme probe span.

Response times: patients vs. controls. The control range for the phonological relatedness by trial type effect was -6 ms to 262 ms (*M*=68, *SD*=89). Unexpectedly, all the patients showed a pattern opposite that of the controls, with a larger phonological relatedness by trial type effect in the *after* condition than the *during* condition. Accordingly, the effects for the patients were negative. For patients DW, LC, PP and RI, the values were -32 ms, -36 ms, -103 ms and -101 ms, respectively. All patients performed outside of the range of controls for the phonological relatedness by trial type effect and performance was not closely related to rhyme or category probe span (see Figure 9a and 9b). Table 14 contains a summary of the data for Experiment 3.

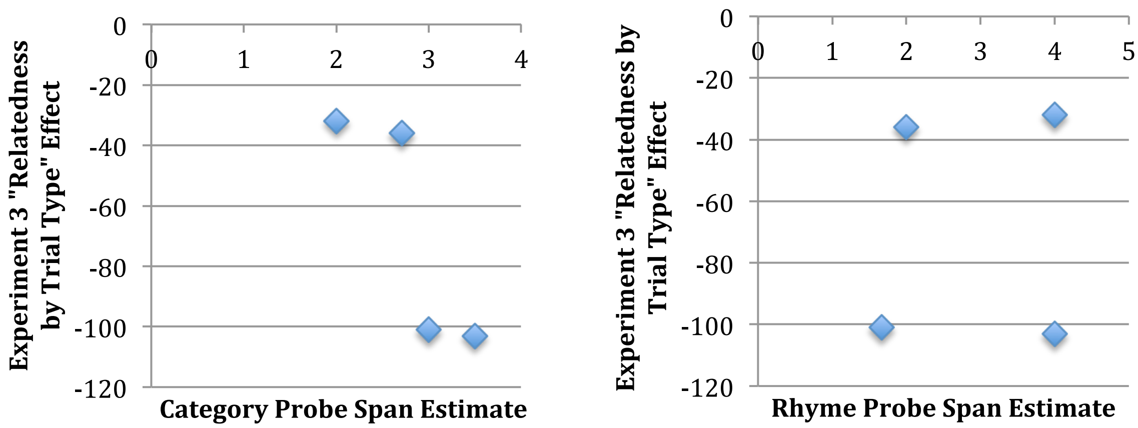

Figure 9. Relationship between probe span and interactions of relatedness by *during* vs. *after* for patients in Experiment 3. Figure 9A (left) presents the relationship between the category probe span and interactions of relatedness by *during* vs. *after* for patients in Experiment 3. Figure 9B (right) presents the relationship between the rhyme probe span and interactions of relatedness by *during* vs. *after* for patients in Experiment 3. As can be seen, performance on Experiment 3 was more closely related to category probe span, not rhyme probe span.

Table 14

Average response times (ms) for lexical decision to probe items by relatedness and trial type, Experiment 3

Relatedness	Phonological, *During*	Unrelated, *During*	Phonological-Unrelated, *During*	Phonological, *After*	Unrelated, *After*	Phonological-Unrelated, *After*
Control Average	868	809	59*	764	773	-9
DW	1074	1123	-49	1138	1155	-17
LC	869	858	11	887	840	47
PP	1016	1078	-62	831	790	41
RI	1697	1669	28	1599	1470	129

Note. Semantic-unrelated refers to the difference between the average response time for semantic and unrelated probe items. A * symbol represents that the difference was statistically significant ($p < .05$) in independent samples t-tests, for patients, or paired-samples t-tests by items, for controls. A + symbol represents that the difference was statistically significant ($p < .05$) in paired-samples t-tests by subjects.

Response times: patients. Response time data were included and analyzed based on the same criteria as Experiment 1A. As in Experiments 1A and 2, the patient response time data were initially analyzed using a 2x2 ANOVA with relatedness (phonologically related vs. unrelated) as a within-items variable and trial type (*during* vs. *after*) as a between-items variable. None of the interaction effects were significant for any of the patients. This might be due to the large number of errors, so independent-samples t-tests were used to analyze the data to increase power in detecting significant effects. The means are shown in Table 13. None of the patients showed any significant effects in the *during* trial type (DW, $p = .5$; LC, $p = .67$; PP, $p = .44$; RI, $p = .78$) or the *after* trial type (DW, $p = .85$; LC, $p = .1$; PP, $p = .47$; RI, $p = .24$).

Discussion: Experiment 3

The findings with age-matched control subjects in Experiment 3 replicated Shivde and Anderson's (2011) finding of a phonological relatedness effect in younger adults. Our age-matched control subjects took longer to respond to phonologically related words compared to unrelated words in the *during*, but not *after*, trial type. Combined with Experiment 1A, this provides further evidence for a dissociation between semantic and phonological maintenance abilities. Additionally, these findings will allow us to draw conclusions regarding our patient performance on this task given that the processes seen in younger adults seem to be in effect in older adults.

As for patient performance on Experiment 3, none of our patients exhibited a phonological relatedness effect or showed any significant effects. Additionally, when examining the interactions of relatedness by *during* vs. *after*, all patients fell outside of the range of controls. If we look back to Experiment 2, we see that the patient results of Experiment 3 all follow the same pattern between their *during* and *after* trials (i.e., they all have negative interactions of relatedness by *during* vs. *after* indicating that the difference between phonologically related and unrelated probe items was larger in the *after* trial type than the *during* trial type) as they did in Experiment 2, reflecting some common maintenance strategy between the two experiments. But, when comparing patient performance on Experiment 3 to the probe tasks, performance was more closely related to their category probe span than their rhyme probe span. This is the opposite of Experiment 2 and the opposite of what we predicted if patients were using phonological maintenance strategies. Thus, as in Experiment 2, it appears that subjects may be using some combination of semantic and phonological maintenance or only semantic maintenance. It is also possible that Experiment 3 is not tapping the same phonological buffer as the rhyme probe span. Martin, Lesch and Bartha (1999) proposed that phonological STM could be further divided into input and output phonology. The rhyme probe task is believed to reflect input phonology, but the concurrent probe task may be tapping output phonology since subjects may be repeating the maintenance word to themselves in order to remember the word (e.g., see Baddeley et al, 1984). So, we compared patient performance on Experiment 3 to the Philadelphia List Repetition Task. In this task, patients are presented with a list of words that they are required to repeat back to the experimenter. On this task, LC is most impaired (i.e., he makes 39 phonological errors, 32.5% of trials, on 2 item repetition), followed by PP (i.e., he makes 17 phonological errors, 14% of trials, on 2 item repetition), RI (i.e., he makes 14 phonological errors, 11.7% of trials, on 2 item repetition), and DW (i.e., she makes 7 phonological errors, 5.8% of trials, on 2 item repetition). However, if we compare the data to patient performance on Experiments 3, there is no clear relationship. Thus, it seems that phonological STM deficits manifest differently in the concurrent probe task than do semantic STM deficits and that the task used in Experiment 3 does not appear to be useful for evaluating phonological STM deficits. It is possible, however, that some revised version of this task might be useful.

Finally, patients may not be actively maintaining the phonological form of the maintenance item to the same degree as controls while at the same time maintaining it in a state that is active enough to make a synonymy judgment (Experiment 2) or a pseudohomophone judgment (Experiment 3). These possibilities for the interpretation of the results of Experiment 2 and Experiment 3 will be addressed further in the General Discussion.

General Discussion

The purpose of the current series of experiments was to evaluate semantic and phonological STM deficits in aphasic patients using a novel concurrent probe paradigm (Shivde & Anderson, 2011). Primarily, we were interested in providing converging evidence for semantic maintenance and semantic STM deficits from the concurrent probe paradigm in evaluating semantic compared to phonological STM deficits given the relative scarcity of tasks designed to tap semantic STM. Experiment 1 was aimed at in providing converging evidence for detecting semantic STM deficits. We specifically compared patient performance on the concurrent probe paradigm to a standard task used to evaluate semantic STM, the category probe task. We found that patients with impaired semantic STM, as indexed by performance on the category probe task, did not exhibit the semantic relatedness effects as healthy undergraduates (Shivde & Anderson, 2011), older adults (Experiment 1 of this study) or patients without semantic STM deficits (Experiment 1 of this study). Thus, it seemed that the concurrent probe task used in Experiment 1 is tapping the same underlying deficit in semantic STM. This is an interesting finding in that the two tasks are designed quite differently, with one task requiring maintenance of several items and immediate recognition of a categorical relation between the probe and the list items and the other requiring the maintenance of a single item over several intervening items in a secondary task. Despite these differences, they seem to be tapping into the same semantic maintenance ability.

Additionally, the results of Experiment 1A provided insight as to whether semantic STM deficits were caused by difficulty inhibiting related information (Hamilton & Martin, 2005) or overly rapid decay (Martin & Lesch, 1996). Patients with semantic STM deficits performed in a manner that is more consistent with the overly rapid decay hypothesis; namely, they did not show any significant difference in the amount of time it took to respond to semantically related compared to unrelated probe words in *during* trial types.

In Experiment 2, we instructed patients to use semantic maintenance strategies, but manipulated the lexical decision probe item to be phonologically related instead of semantically related to the maintenance item. We believed that this would be useful in establishing or refuting the recruitment of a phonological maintenance strategy by our patients with deficits in semantic STM. Given the finding of a phonological relatedness effect by Shivde and Anderson (2011) in phonological maintenance conditions, the pattern of results in Experiment 2 for our age-matched control subjects and all of our patients did not support the use of a phonological maintenance strategy. That is, given the lack of phonological relatedness effects for our control subjects and patients, we conclude that they were not using phonological maintenance strategies in semantic maintenance conditions.

However, the pattern of performance for our patients in the task of Experiment 2 was difficult to interpret. Patient RI did not exhibit any significant effects, which would be expected given his relatively intact semantic STM and impaired phonological STM but patient PP, who also has intact semantic STM, showed significantly faster RT's for phonologically related probes in *during* trial types. Further, DW and LC both showed significantly slower RT's for phonologically related than unrelated probes in *after* trial types. It was not clear why this would be the case. Thus, in an attempt to resolve the questions left after Experiment 2, Experiment 3 was administered to the patients to evaluate phonological maintenance strategies using a task that explicitly instructed subjects to use phonological maintenance.

In Experiment 3, older adult controls replicated the findings of Shivde and Anderson (2011) with undergraduates, namely they exhibited a phonological relatedness effect, further supporting the dissociation between semantic and phonological STM that has been observed in previous studies (e.g., Barde et al, 2010; Martin & Lesch, 1996; Martin, Shelton & Yaffee, 1999); however, the patient results did not follow the same pattern as controls. All patients performed outside the range of controls on this task but showed no significant effects. Further, their performance did not seem to be related to their phonological STM span. As mentioned in the discussion of Experiment 3, it is possible that the concurrent probe task of Experiment 3 is not tapping the same phonological STM buffer as the rhyme probe task (i.e., input phonological STM). Given that the rhyme probe task is believed to tap the input phonological buffer, it is possible that the concurrent probe task is tapping the output phonological buffer, explaining why patient performance in Experiment 3 was not related to rhyme probe span and possibly reflecting a rehearsal process involved in maintaining the maintenance item during the lexical decision trials. To evaluate this hypothesis, we compared performance on Experiment 3 with a task known to tap output phonological STM, the Philadelphia List Repetition Task. However, when we compared patient performance on the Philadelphia List Repetition Task to performance on the concurrent probe task, no relationship was found. Thus, it does not appear that output phonology was the factor affecting performance on Experiment 3.

A further possibility is that the patient data from Experiments 2 and 3 were simply too noisy to detect any effects. If this were the case, then the significant effects could have been spurious and non-significant effects could be significant with more data. Further, patient performance between the two experiments followed the same pattern. That is, for each patient, they showed a negative relatedness by trial type interaction effect in Experiment 2 and in Experiment 3, suggesting that the two tasks share some common factor. Regardless, it is clear that the concurrent probe task as designed in Experiment 3 is not useful for evaluating phonological STM deficits.

Finally, it is possible that orthographic presentation of the task did not necessitate the use of phonological maintenance in patients and may have allowed patients to use orthographic maintenance strategies. It is possible that a task with auditorily presented items would provide a better test, as phonological maintenance is obligatory with auditory input whereas it may be optional with written input. Additionally, in Experiments 2 and 3, the phonologically related probe items were not controlled for orthographic similarity to the maintenance or target items, and in Experiment 3 there was no attempt to control for orthographic similarity between the maintenance item and the pseudohomophone target item. Accordingly, many of the probe items in Experiment 2 and 3 and the target items in Experiment 3 were orthographically similar to the maintenance items, which may have confounded the results and contributed to the unexpected findings in patients. Future phonological versions of the concurrent probe task should be sure to control for orthographic similarity and, if possible, attempts should be made to create an auditory version of this task.

Future research will focus on testing more patients on the task of Experiment 1A to provide more evidence for the usefulness of this measure in detecting semantic STM deficits. It will be important to replicate the findings in Experiment 1 with a larger group of patients and, with a larger group of patients, we will be able to perform correlation analyses comparing category probe span and performance on the concurrent probe task and thus further evaluate the relationship between these two tasks. Based on the findings from Experiment 1A, it can be predicted that a correlation analysis will reveal a high correlation between the two tasks. Finally, future research should look at the distinction between inhibition deficits and overly rapid decay hypotheses in relation to semantic STM deficits using other paradigms. Specifically, it will be important to find converging evidence for the overly rapid decay hypothesis. Other researchers have looked closely at the relationship between inhibition and semantic STM and found no significant correlations between these two types of task (e.g., Allen, Martin & Martin, 2012), so following up on the overly rapid decay hypothesis may reveal the source of semantic STM deficits.

In conclusion, given that there are many tasks useful in evaluating phonological STM (e.g., rhyme probe, digit span, digit matching), it was much more important to show that the concurrent probe task provides converging evidence for semantic STM deficits, which is generally measured using only one task (i.e., the category probe task). Given the findings of Experiment 1A-1C, it seems clear that the concurrent probe task is useful in detecting semantic retention deficits but, at least in its current form, not in detecting phonological retention deficits.

References

Allen, C., Martin, R., & Martin, N. (2012). Relations between short-term memory deficits, semantic processing, and executive function. *Aphasiology, 26*, 428-461.

Baddeley, A.D., & Hitch, G.J. (1974). Working memory. In G.A. Bower (Ed.), *The Psychology of Learning* (pp. 47-89). London: Academic Press.

Baddeley, A.D., Lewis, V., & Vallar, G. (1984). Exploring the articulatory loop. *The Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology, 36*(2), 233-252.

Barnard, P. (1985). Interacting cognitive subsystem: A psycholinguistic approach to short-term memory. In A.W. Ellis (Ed.), *Progress in the psychology of language,* Vol. 2 (pp. 197-258). London: Lawrence Erlbaum Associates.

Caplan, D., & Waters, G. S. (1999). Verbal working memory and sentence comprehension. *Behavioral and Brain Sciences*, *22*, 77-126.

Cohen, J. D., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: A new graphic interface environment for designing psychology experiments. *Behavioral Research Methods, Instruments, and Computers, 25*(2), 257-271.

Collins, A., & Loftus, E. (1975). A spreading-activation theory of semantic processing. *Psychological Review, 82*(6), 407-428.

Crawford, J.R., & Howell, D.C. (1998). Comparing an individual's test score against norms derived from small samples. *Clinical Neuropsychologist, 12*, 482-486.

Crowder, R.G. (1978). Memory for phonologically uniform lists. *Journal of Verbal Learning and Verbal Behavior, 17*, 73–89.

Freedman, M., & Martin, R. (2001). Dissociable components of short-term memory and their relation to long-term learning. *Cognitive Neuropsychology*, 18, 193-226.

Glanzer, M. (1969). Distance between related words in free recall: Trace of the STS. *Journal of Verbal Learning and Verbal Behavior, 8*, 105-111.

Haarman, H., & Usher, M. (2001). Maintenance of semantic information in capacity-limited short-term memory. *Psychonomic Bulletin & Review, 8*(3), 568-578.

Hamilton, A.C., & Martin, R.C. (2005). Dissociations among tasks involving inhibition: A single-case study. *Cognitive, Affective, & Behavioral Neuroscience, 5*(1), 1-13.

Hamilton, A.C., & Martin, R.C. (2007). Proactive interference in a semantic short-term memory deficit: Role of semantic and phonological relatedness. *Cortex, 43*, 112-123.

Hamilton, A.C., Martin, R.C., & Burton, P.C. (2010). Converging functional magnetic resonance imaging evidence for a role of the left inferior frontal lobe in semantic retention during language comprehension. *Cognitive Neuropsychology, 26*(8), 685-704.

Howard, D., & Patterson, K. E. (1992). *The Pyramids and Palm Trees Test: A test of semantic access from words and pictures*. Thames Valley Test Company.

Hulme, C., Maughan, S., & Brown, G. (1991). Memory for familiar and unfamiliar words: Evidence for a long-term memory contribution to short-term span. *Journal of Memory and Language, 30*, 685–701.

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

Kay, J., Lesser, R., & Coltheart, M. (1992). *PALPA: Psycholinguistic assessments of language processing in aphasia*. Psychology Press.

Kucera, H., & Francis, W. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.

Landauer, T.K., Foltz, P.W., & Laham, D. (1998). Introduction to latent semantic analysis. *Discourse Processes, 25*, 259-284.

Majerus, S., Van Der Linden, M., & Renard, C. (2001). Short-term memory and language processing: Further evidence for the existence of separate phonological and semantic short-term memory components. *Current Psychology Letters: Brain Behavior & Cogntion, 4*, 23-37.

Martin, R.C., Crowther, J.E., Knight, M., Tamborello II, F.P., Yang, C. (2010). Planning in sentence production: Evidence for the phrase as a default planning scope. *Cognition, 116*, 177-192.

Martin, R. C., & Freedman, M. L. (2001). Short-term retention of lexical-semantic representations: Implications for speech production. *Memory, 9*, 261- 280.

Martin, R.C., & He, T. (2004). Semantic short-term memory and its role in sentence processing: A replication. *Brain and Language, 89*, 76-82.

Martin, R. C., & Lesch, M. F. (1996). Associations and dissociations between language impairment and list recall: Implications for models of short-term memory. In S. Gathercole (Ed.), *Models of short- term memory* (pp. 149-178). Hove, U.K.: Erlbaum.

Martin, R.C., Lesch, M.F., & Bartha, M.C. (1999). Independence of input and output phonology in word processing and short-term memory. *Journal of Memory and Language, 41*, 3-29.

Martin, R. C., Miller, M., & Vu, H. (2004). Lexical-semantic retention and speech production: Further evidence from normal and brain-damaged participants for a phrasal scope of planning. *Cognitive Neuropsychology*, *21*, 625-644.

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

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

Shivde, G., & Anderson, M.C. (2011). On the existence of semantic working memory: Evidence for direct semantic maintenance. *Journal of Experimental Psychology: Learning, Memory and Cognition, 37*, 1342-1370.

Shivde, G., & Thompson-Schill, S. (2004). Dissociating semantic and phonological maintenance using fMRI. *Cognitive, Affective & Behavioral Neuroscience, 4*(1), 10-19.

Wilson, M.D. (1988). The MRC Psycholinguistic Database: Machine Readable Dictionary, Version 2. *Behavioural Research Methods, Instruments and Computers*, *20* (1), 6-11.

Appendix A

Experiment 1 Materials

Maintenance Item (semantic maintenance)	Probe Item (semantically related)	Target Item (synonym)
afraid	snake	fearful
angry	fight	mad
answer	question	response
bad	cheat	corrupt
brag	pride	boast
build	castle	erect
busy	tired	occupied
cautious	danger	wary
center	circle	middle
change	clothes	alter
cheap	sale	inexpensive
cold	wind	freezing
correct	response	right
courage	hero	bravery
dead	sick	deceased
different	alike	contrary
disgrace	embarrass	shame
display	show	exhibit
easy	quick	simple
embrace	arms	hug
emotion	cry	feeling
endure	problem	survive
enjoy	fun	like
equal	rights	even
exit	enter	leave
fat	thin	plump
fever	sick	temperature
find	lose	locate
first	last	primary
fraud	fake	deceit
freedom	slavery	liberty
genuine	real	authentic
glad	satisfied	happy
go	race	proceed

Maintenance Item (semantic maintenance)	Probe Item (semantically related)	Target Item (synonym)
grasp	release	clutch
happen	event	occur
hate	love	disdain
help	trouble	aid
hike	road	trek
honest	liar	truthful
honor	dignity	glory
hurt	harm	pain
idea	opinion	thought
impostor	disguise	pretender
long	short	lengthy
mankind	world	humanity
mistake	sorry	error
nothing	empty	zero
old	new	aged
outcome	win	result
part	portion	piece
permit	deny	allow
pity	mercy	sympathy
prank	fool	joke
present	past	current
purchase	price	buy
quarrel	dispute	spat
rascal	mischief	scoundrel
reach	goal	extend
reject	accept	renounce
rejoice	pleasure	celebration
rise	fall	ascend
same	similar	alike
send	receive	transmit
shape	square	labor
shiny	silver	glossy
skill	clever	talent
sleep	dream	slumber
sob	tears	weep
sorrow	misery	grief
source	beginning	origin
speak	tongue	converse

Maintenance Item (semantic maintenance)	Probe Item (semantically related)	Target Item (synonym)
stamina	sport	energy
stop	sign	quit
strong	big	muscular
succeed	famous	achieve
test	score	exam
think	brain	ponder
tired	nap	fatigued
tour	visit	expedition
triumph	overcome	victory
walk	run	stroll
wander	aimless	roam
want	crave	desire
warm	summer	hot
weak	feeble	frail
wild	animal	savage
work	job	labor
young	child	adolescent

Appendix B

Experiment 2 Materials

Maintenance Item (semantic maintenance)	Probe Item (phonologically related)	Target Item (synonym)
abundant	funded	numerous
accomplish	dish	achieve
acquire	tire	obtain
approach	roach	arrive
argue	dew	fight
assume	bloom	infer
below	slow	beneath
blame	name	fault
bliss	swiss	joy
book	look	novel
boring	scoring	dull
choose	lose	select
clever	lever	smart
collect	respect	gather
conflict	predict	clash
constant	vent	fixed
crash	flash	accident
crime	lime	theft
crook	hook	thief
crooked	wicked	bent
cut	shut	slice
damp	cramp	moist
decide	ride	determine
defeat	meat	beat
definite	senate	certain
delicious	vicious	savory
deny	fly	prohibit
describe	tribe	portray
destroy	toy	ruin
difficult	insult	hard
distress	dress	danger
excess	press	surplus
explain	rain	clarify
fall	ball	drop

Maintenance Item (semantic maintenance)	Probe Item (phonologically related)	Target Item (synonym)
fast	last	quick
fight	bite	brawl
fly	try	soar
funny	money	silly
garbage	porridge	trash
give	live	donate
gone	dawn	absent
hideous	insidious	ugly
honest	fawn	fair
hurry	jury	rush
ill	fill	sick
impartial	skull	neutral
infant	instant	baby
informal	normal	casual
insert	flirt	install
job	bob	duty
jog	dog	run
kill	skill	murder
leap	steep	jump
lie	die	deceive
listen	chicken	hear
look	shook	gaze
mend	bend	repair
nasty	dynasty	gross
near	beer	close
nervous	service	anxious
pathetic	athletic	pitiful
perfect	eject	flawless
plan	van	plot
pledge	edge	oath
postpone	bone	delay
problem	bum	issue
prove	move	verify
pull	full	tug
raise	praise	lift
regular	settler	routine
rent	sent	lease

Maintenance Item (semantic maintenance)	Probe Item (phonologically related)	Target Item (synonym)
rest	test	relax
resume	boom	return
risk	brisk	gamble
sack	tack	bag
scheme	theme	plot
shout	doubt	yell
shut	nut	close
slow	glow	sluggish
sniff	cliff	inhale
sort	court	classify
strange	change	odd
tardy	hardy	late
terrible	wool	awful
threat	bet	risk
tiny	shiny	small
understand	grand	comprehend
untrue	blue	false
use	fuse	employ
weight	bait	mass

Appendix C

Experiment 3 Materials

Maintenance Item (phonological maintenance)	Probe Item (phonologically related)	Target Item (pseudohomophone)
binge	hinge	binje
bird	herd	berd
birth	worth	burth
blood	mud	blud
boat	note	bote
bone	stone	boan
box	rocks	bawx
breath	death	breth
burn	return	bern
certain	curtain	sertain
cheat	greet	cheet
chief	beef	cheef
choice	voice	choyce
code	load	coad
concept	slept	consept
cough	off	coph
date	late	dait
dead	bed	ded
dirt	shirt	durt
doll	call	daul
doubt	shout	dowt
dream	team	dreem
feel	steal	feal
firm	perm	ferm
first	worst	furst
foal	goal	fole
foam	chrome	phoam
found	mound	fownd
fox	locks	phocks
gain	lane	gane
ghoul	pool	gool
girl	twirl	gerl
glue	chew	glew

Maintenance Item (phonological maintenance)	Probe Item (phonologically related)	Target Item (pseudohomophone)
green	lean	grean
group	stoop	groop
grow	low	groe
guide	hide	gide
has	jazz	hazz
hope	cope	hoap
keys	knees	keeze
kite	bite	kight
learn	turn	lern
lease	geese	leese
least	beast	leest
mate	bait	mait
mean	bean	meen
meant	lint	ment
money	funny	munny
mule	fuel	muel
murder	herder	merder
news	lose	nuez
nurse	purse	nerse
once	dunce	wunce
pearl	whirl	pirl
perk	clerk	pirck
piece	fleece	peece
proof	goof	prufe
prove	move	pruve
purge	surge	perge
raise	gaze	raize
reach	teach	reech
role	dole	roal
roof	aloof	rufe
serve	nerve	surve
shoes	clues	shooze
slight	tight	slite
snow	glow	snoe
soap	rope	sope
soon	noon	sune
soup	scoop	supe
stop	hop	stawp

Maintenance Item (phonological maintenance)	Probe Item (phonologically related)	Target Item (pseudohomophone)
suit	boot	sute
sure	cure	shure
tape	grape	taip
term	worm	turm
threat	bet	thret
tracks	jacks	trax
wage	page	waije
wake	take	waik
wall	fall	wawl
weed	need	wead
white	sight	wite
wife	life	whife
word	heard	werd
work	jerk	werk
worth	earth	wirth
wrong	strong	rong
year	deer	yeer
young	hung	yung

Appendix D

Post-Experiment Survey

Post-Experiment Survey: Concurrent Probe Paradigm

1. What strategy did you use to maintain the red word?

2. Did you change your strategy when the red and blue words were farther apart?

3. What were your perceptions of the task? Did you find it difficult?