Multiple Memory Systems?:
Serial Position Dependent False Memory Effects

by

Halle R. Zucker

A Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of Bachelor of Arts
With Honors in Brain, Behavior, and Cognitive Science from the
University of Michigan
2010

Advisor: Dr. Patricia A. Reuter-Lorenz

Abstract

The century-long debate about memory’s structure continues today based on behavioral and neuroscience data. To investigate the dissociability of short- and long-term memory (S/LTM) we combined the classic logic of the serial position curve with the Deese-Roediger-McDermott (DRM) false memory task. Participants studied 12-item lists comprised of 3 sublists containing 4 strong associates of a non-presented theme word. Experiment 1 used immediate free recall whereas a 3-second filled retention interval preceded recall in Experiment 2. Both experiments showed false recall associated with primacy and recency positions. The recency effect on recall and phonological errors associated with recency positions was diminished in Experiment 2. The results suggest that semantic processes operate in S/LTM and differ in prominence depending on interference.

Multiple Memory Systems?: Serial Position Dependent False Memory Effects

The Structure of Memory

Memory provides a foundation for all cognitive and behavioral processes. It allows us to have a sense of self and to function in the day-to-day world. At least since the classic works of Ebbinghaus (1885), psychologists have been investigating how memory is structured. At the heart of this investigation is the question of whether the same memory system mediates our ability to remember a new phone number long enough to dial it (short-term retention) and our ability to enduringly remember our own number (long-term retention). The goal of this project is to provide new evidence pertaining to the structure of memory by testing the influence of semantic processing in putatively separate memory systems.

A time-honored method used to address this question is serial list learning (Murdock, 1962; for a recent review see Laming, 2010). In such a task, participants study lists of 12-15 items and then their memory for the list is tested via free recall—recalling the words without specifications on output order. Typically, the first few and last few items are remembered better whereas memory for items presented near the middle of the list is impoverished (Deese & Kaufman, 1957). Higher performance for early items, the primacy effect, has been associated with storage in long-term memory (LTM) whereas higher performance for later items, the recency effect, has been associated with storage in a separable short-term memory system (STM; Atkinson & Shiffrin, 1968). The probability of recalling these items as a function of presentation position is described graphically in the serial position curve (e.g., Murdock, 1962). Converging evidence that the primacy and recency portions of the serial position curve tap distinct short- and long-term memory (S/LTM) systems—with separable encoding, storage, and retrieval processes—comes from differential performance across list positions in amnesics (Milner, 1978), differential impacts of interleaved time on forgetting (Glanzer & Cunitz, 1966; Postman & Phillips, 1965), differential distraction effects across list positions (Baddeley, 2003, p.32), and differential rates of semantic and phonological errors across serial positions (Vallar & Shallice, 1990, p. 21-22).

Multiple-store models of memory have been founded on evidence for dissociations—finding that some factor affects one type of memory but not the other—across a range of tasks, subject populations, and methodologies. Behavioral work (Glanzer & Cunitz, 1966) and neuroimaging studies (Talmi, Grady, Goshen-Gottstein, & Moscovitch, 2005), as well as data from neuropsychological patients (Baddeley & Warrington, 1970; Scoville & Milner, 1957; Shallice & Warrington, 1970), support the notion of separate stores for S/LTM. Long-term memory is thought to rely on semantic, or meaning-based codes (Baddeley, 1966), that can hold infinite amounts of material for indefinite periods of time (Bahrick, 1984). On the other hand, STM is thought to rely on phonological, or surface-based codes (Sperling, 1960; Wickelgren, 1965) with a limited capacity (Cowan, 2000; Miller, 1956).

Recent evidence has begun to shift the tide against multiple-store models in favor of a unitary memory system with a single store. While the concept of a unitary store is not novel (e.g., Melton, 1963), recent neuroscience evidence to corroborate this idea comes from patient evidence that neural structures implicated in LTM storage are recruited at short delays (Hannula, Tranel, & Cohen, 2006) and neural evidence that common regions are activated during retrieval from S/LTM (Cabeza, Dolcos, Graham, &

Nyberg, 2002; Ranganath & Blumenfeld, 2005). These accounts posit that memory over the short- and long-term share the same representational bases and are subject to similarity-based interference (e.g., Jonides et al., 2008). Unitary models generally posit that STM is simply the activated portion of LTM (Cowan, 2000).

False Memories in Long-Term Memory

An important feature of memory, that may hold important clues to its structure, is when and how it fails us. One type of memory error that has received major research focus is “false memory” (for a comprehensive review see Gallo, 2006). False memories occur when someone misremembers the details of an event or remembers an event that did not happen at all. These false memories are spontaneous, unintentional experiences and are inherently different from outright lies or omissions (Schacter, 1999).

This phenomenon first gained widespread public attention with accusations of therapists recovering repressed memories (Alpert et al., 1998; Loftus, 1997). Research has verified the malleability of memory showing that participants will falsely recall being lost in a mall as a child (Loftus & Pickrell, 1995), nearly drowning (Heaps & Nash, 2001), or spilling punch on the bridal party at a wedding (Hyman, Husband, & Billings, 1995). Such results provide evidence that autobiographical false memories can be reliably produced and give credence to the claim that we can misremember details of our life experience.

A popular and efficient method for investigating false memories in the lab is the Deese-Roediger-McDermott (DRM) task (Deese, 1959; Roediger & McDermott, 1995). In this task, participants are presented with lists of 12 items that are related in meaning to a non-presented theme word (e.g., *bed, rest, awake, tired, dream, wake, snooze, blanket,* *doze, slumber*, *snore, nap, peace, yawn*, and *drowsy* that are all related to the non-presented theme word *sleep*). Participants are then asked to either recall as many of the studied items as possible from memory or recognize studied items from lists containing both presented and non-presented items.

Surprisingly, participants frequently both falsely recall and falsely recognize non-presented themes or other semantic associates. These false memories can occur at similar levels of confidence to veridical memories (Payne, Elie, Blackwell, & Neuschatz, 1996) and even after preventative warnings (Gallo, Roediger, & McDermott, 2001) or strategy coaching (Anastasi, Rhodes, & Burns, 2000).

These false memory errors fall within the domain of LTM because the lists typically exceed estimates of STM capacity (i.e., approximately 4 items; Cowan, 2000) and they are also tested across longer durations ranging from minutes (Roediger & McDermott, 1995) to months (Seamon et al., 2002). Because LTM is thought to rely on semantic codes (Baddeley, 1966) meaning-based distortions, as well as the fact that these tasks are designed to tap into LTM, are taken as evidence that these responses are generated from the long-term store.

Many theories have attempted to explain why these false memory errors occur including associative activation and thematic consistency (see Gallo, 2006 for a review). Associative activation assumes the existence of an internal mental lexicon with a semantic, node-based organization. Activation of one concept will cause other, related nodes to be activated and can predict false memory performance (Roediger, Watson, McDermott, & Gallo, 2001). This activation may even occur implicitly at the time of study (Underwood, 1965). In contrast, thematic consistency assumes that what causes false memories is a similar underlying semantic theme. It conceptualizes two traces, verbatim and gist, which compose the overall semantic theme. Verbatim traces encode surface forms and contextual cues which support detailed, accurate recall whereas gist traces represent overall concepts such as meanings and overarching patterns which support memory for an overall scene or idea. The item-specific verbatim traces fade more quickly than the concept-rich gist traces. Veridical memories are supported by both verbatim and gist traces whereas false memories occur when gist traces are present in the absence of verbatim traces (Brainerd & Reyna, 2002). Both of these theoretical explanations also call on the tradition of LTM since they are based on semantic features of the studied items as an explanation of later memory errors.

False Memories in Short-Term Memory

According to the traditional view of separable systems, STM should be relatively invulnerable to the semantic distortions found in LTM. Short-term memory is thought to rely on perceptually-based (e.g., phonological, orthographic) codes (Sperling, 1960; Wickelgren, 1965) and only minimally on the deeper, semantic codes of LTM.

Atkins and Reuter-Lorenz (2008) provide evidence that challenges this view in a study that adapted the DRM task to examine semantic distortions in STM. They presented participants with four-item lists constructed by trimming the 12-item DRM lists to their four highest associates (e.g. *nap, doze, bed,* and *awake* would comprise the STM *sleep* list) as measured by backward associative strength (BAS)—a measure of how likely the theme words are to elicit the studied cues (e.g. how likely one is to generate “haystack” when given the word “needle”). Lists were probed a mere three-to-four seconds following their presentation. Across both recognition and free recall, participants falsely remembered lure items and other semantic associates. These results indicate that semantic false memories are not exclusive to LTM and, along with other evidence discussed above (Cabeza et al., 2002; Cowan, 2000; Hannula et al., 2006; Jonides et al., 2008; Melton, 1963; Ranganath & Blumenfeld, 2005), call into question sharp distinctions between S/LTM systems.

Current Study

This study uses the logic of the serial position curve (Murdock, 1962) in conjunction with the DRM false memory task (Deese, 1959; Roediger & McDermott, 1995) to investigate false memories across S/LTM in a single subject population. Here, we were interested in the systematic occurrence of errors across presentation positions within a studied list.

Our strategy was to test serial recall of a 12-item list composed of three semantically distinct sublists (e.g., *absent, gift, future, past, circle, round, triangle, shape, keep, grip, grasp,* and *carry* where the three unique semantic themes were *present, square,* and *hold*). Each sublist contained four words which were semantic associates of the same theme word, although the theme words were never presented. At retrieval, participants freely recalled items from memory which permitted the derivation of serial position curves. Critically, by measuring rates of errors associated with each sublist, false memories in the recency versus primacy portion of the curve could be compared.

We seek to determine if semantic errors will occur in the recency portion of a list of memoranda at recall. Multiple-store models would predict that semantic false memories from the recency (STM) portion of the list should be rare. Unitary-store models, and recent evidence from Atkins and Reuter-Lorenz (2008), would predict that the semantic (gist) codes present in the recency portion would yield robust semantic false memories coupled with high veridical memories. We hypothesize that we will find semantic false memories from both primacy (LTM) and recency (STM) portions of the list. Furthermore, we expect that including a filled retention interval prior to recall will decrease the recency effect in veridical recall while increasing the number of semantic errors from this portion of the list.

Experiment 1 investigates false memories in immediate free recall. This methodology is extended in Experiment 2 which inserts a filled retention interval prior to recall.

Experiment 1: Investigations in False Memory with Free Recall

Method

Participants. Fifty-two University of Michigan undergraduate students took part in the experiment for either course credit or monetary compensation ($10 per hour). All participants were native English speakers, right-handed, and free from any reported neurological or psychological conditions. Three participants were excluded due to experimenter error and equipment malfunctions, four participants were excluded for audibly reading the studied words, two participants were excluded for responding strategies that significantly changed the time demands of the task (perseverative repetition of list items [more than four repetitions per trial], a severe speech impediment), and one participant was excluded for being on psychoactive medications. The following analyses represent data from the remaining 42 participants (14 female; *M* age = 19.45 years, *SD* = 1.12).

Materials. One 12-item list was presented on each of 42 trials. Each list was composed of three quartets where each quartet contained four words that were all semantically related to one common non-presented theme word (see Appendix). The quartets in each list were counterbalanced so that each triplet appeared equally as often in the primacy (“A”), middle (“B”), or recency (“C”) position. Additionally, the order of items within each quartet was randomized across subjects although semantic grouping across list positions was preserved.

Quartets represent a subset of 126 lists chosen from 136 lists consisting of four words converging on a common semantic associate taken from previously published lists used to elicit false memories in S/LTM (Flegal, Atkins, & Reuter-Lorenz, in press). Quartets were semantically distinct lists defined by mean BAS (*M* BAS = 0.34, *SD* = 0.16). Within each 12-item list, each quartet was categorized as either the high, medium, or low BAS list. This classification system was relative within each list and there were no standardized cut-offs for each BAS classification. Each BAS categorization (“high”, “medium”, “low”) served equally as often in each sublist position (“A”, “B”, “C”). There was no effect of BAS position so all analyses presented below are collapsed across this variable.

Procedure. Participants first gave written consent and then completed the WAIS-III Digit Span (Wechsler, 1997) task before proceeding to the testing room. Following this, participants completed the free recall task on the computer and then a pencil and paper recognition task following completion of all recall trials. A computerized version of the operation span task (Unsworth, Heitz, Schrock, & Engle, 2005) was given to obtain estimates of working memory span. A modified version of a source memory task (Drag, Bieliauskas, Kaszniak, Bohnen, & Glisky, 2009) was used to assess vulnerability to source memory errors. All participants filled out a post-test survey and were debriefed. Means for digit span, operation span, and source memory tasks are presented in Table 1. The results of these ancillary tasks will not be discussed further here.

Recall task. Participants were provided with a hard copy of the instructions to accompany those presented on the computer screen. The experimenter read these instructions aloud and answered any questions. The recall task was prefaced with two practice trials to familiarize participants with the structure of the task. Audio recording commenced at the beginning of the first practice trial and continued for the duration of the experiment.

After reviewing the task instructions and completing two practice trials, the participant began the actual experiment. An experimenter remained in the testing room for the entire testing procedure. Words were presented serially at a rate of 800 ms per word (after Talmi et al., 2005) with 50 ms between words. Timing and presentation parameters were controlled using ePrime 2.0 software (Psychology Software Tools, Pittsburgh, PA). Stimuli appeared in a black Arial 17 point font on a silver background. Presentation of the 12 words was followed by a row of five green “X”s for 300 ms that served both as a visual mask as well as the recall cue. Participants were required to take a minimum of 30 seconds to freely recall aloud the studied items. Instructions emphasized that participants should try to recall all 12 items and to recall items they *thought* had appeared even if they were not entirely certain. Auditory responses were collected with Olympus WS-210S digital voice recorders. When the participant felt that they had recalled all of the words in their memory, they pressed the space bar on a standard

QWERTY keyboard to advance to the next trial. Subsequent trials began either 1500 or 2000ms following pressing the space bar.

Data analysis. Two trained coders transcribed each participant's audio responses as either correct or incorrect. A third coder was used to resolve discrepancies in coding classifications. Incorrect responses were further classified by error type within each sublist ("A", "B", or "C") as follows in accordance with previous work (Atkins & Reuter-Lorenz, 2008): (1) s*emantic:* responses judged by both coders as related in meaning to an item in the memory set, or the theme word itself, but not actually presented in the trial, (2) *lure:* reporting the unstudied theme word as one of the studied items, a special case of a semantic error, (3) *phonologic:* related in sound to one of the items in the memory set, (4) *repeated:* an item presented in one of the two previous trials or a correct response that was repeated within a single trial, (5) *unrelated:* not semantically or phonologically related to any items in the memory set and not a member of one of the two previous trials, (6) *mispronunciations:* items where the participant was unable to correctly say the word due to a supposed lack of familiarity but not due to a lack of remembering the studied word, and (7) *non-word utterances:* creation of non-standard English words that were unable to be classified as either phonologic errors or mispronunciations. Response times as recorded in ePrime were also obtained but will not be discussed in this paper. Mispronunciations, repeats, and non-word utterances were not included in any analyses.

Results

Participants correctly recalled an average of 8.23 words per trial (*SD* = 1.07). Proportion of correct recall responses as a function of list position were plotted to derive the classic serial position curves (see Figure 1A). These curves reveal the expected pattern of enhanced recall for the first and last several positions with the lowest recall for the intermediate portions of the list.

Mean correct recall responses by presentation positions (“A”, “B”, “C”; hereafter called sublist) are plotted in Figure 2A. Responses varied reliably across sublist, $F(2,82) = 151.68$, $p < .001$. Participants accurately recalled more items from the “C” sublist than either the “A”, $t\ (41) = -12.57$, $p <.001$, or “B”, $t\ (41) = -16.57$, $p <.001$, sublists. Mean number of correct recall responses did not reliably differ between “A” and “B” sublists, $t\ (41) = 1.64$, $p =.11$.

Out of all errors, 39% were lure items, 26% were semantically related yet non-theme items, 2% were phonologic, and 10% were unrelated items. The remaining 23% of errors were repeated items (both within and between trials) and non-word utterances. Mispronunciations were not included either as errors or correct responses. Participants made an average of 0.03 ($SD = 0.01$) errors per trial. Mean recall errors by error type are plotted in Figure 3A. There was a significant main effect for error type (semantic, phonologic, unrelated) collapsing across sublist, $F(2,82) = 185.13$, $p < .001$. Participants made significantly more semantic errors than either phonologic, $t(41) = 14.61$, $p <.001$, or unrelated errors, $t(41) = 13.12$, $p <.001$. Additionally, they made more unrelated errors than phonologic errors, $t(41) = -5.74$, $p <.001$.

There was a significant interaction between sublist (“A”, “B”, “C”) and error type (semantic, phonologic), $F(2,82) = 45.90$, $p < .001$. Differences between semantic and phonologic errors were significant at every sublist (ps < .001). Means for both semantic and phonologic errors are presented Figure 4A. As can be seen in Figure 4A, semantic errors decreased significantly between the primacy and recency sublists, $t(41) = 8.50$, p

<.001, whereas phonological errors increased significantly between the primacy and recency sublists, *t*(41) = -3.42, *p* =.001. Unrelated errors were not included in this analysis because they cannot be categorized by sublist.

Discussion

Overall, participants performed well on the recall task. They recalled, on average, nearly two-thirds of the 12 items presented on each list.

The results from Experiment 1 show the classic primacy and recency effects associated with positions 1 and 2, and 11 and 12 respectively. In addition, a release from proactive interference effect appears to be superimposed as is evident in recall enhancement in the transitions between sublists (positions 4 vs.5 and 8 vs. 9). Proactive interference is the ability of previous information to negatively impact current performance. Giving out your prior phone number instead of your current one is an example of this effect. A release from proactive interference is traditionally observed when the category of the studied material changes—for instance, names of flowers to names of colors (Underwood, 1945; Wickens, 1970). This shows that the semantic subdivisions were sufficiently salient to produce a release from proactive interference.

The most frequent type of errors made were semantic in nature. Of these errors, theme words (“lures”) were the most common, presumably due to their high normative association with the sublists. This result would be predicted by the associative activation theory because lures should have the closest connections to studied items as represented by normative BAS ratings. Semantic errors were prevalent at all sublists, reaching their highest frequency for the primacy sublist and the lowest for the recency sublist. In contrast, phonological errors were rare but showed a slight, and significant, increase in association with the final sublist. This pattern of errors conforms to the expectations of the multiple-store memory view, according to which semantic and phonologic codes are associated with LTM and STM respectively.

However, these data are not fully compatible with multiple-store models since the presence of semantic distortions in the recency sublist indicates that gist-based coding, and the resultant meaning-based memory errors, are not unique to the primacy sublist. This suggests that there is a continuity of processing across sublists. Despite the relatively strong availability of verbatim memory, which led to high veridical recall, gist-based errors were more likely than any other error types, even from the recency sublist. The increase in phonologic errors from the recency portion compared to earlier sublists is consistent with the greater strength of these codes in immediate recall. But, the fact that there were *any* semantic errors occurring from the recency sublist counters the claim of fully dissociable memory stores for S/LTM.

The presence of semantic distortions from the recency sublist motivates further investigation into the structure of STM in this modified DRM-serial position curve task because it indicates that there may be a similar underlying semantic code shared between S/LTM. In the second experiment, we include a filled retention interval preceding recall in each list. In traditional list learning experiments, inclusion of a filled retention interval should diminish or eradicate the veridical recency effect (Glanzer & Cunitz, 1966; Postman & Phillips, 1965). If this filled retention interval also increases semantic errors from the recency portion, as in Atkins and Reuter-Lorenz (2008), this would be evidence for the operation of similar codes in S/LTM. Additionally, it would show that both gist and verbatim traces are present in STM.

Experiment 2: Investigations in False Memories with Delayed Free Recall

Method

Participants. Fifty-five University of Michigan undergraduate students took part in the experiment for course credit. All participants were native English speakers, right-handed, and free from any reported neurological or psychological conditions. Six participants were excluded due to experimenter error and equipment malfunctions, two participants were excluded for audibly reading words aloud during the study phase, and five participants were excluded for having accuracy on the retention interval math task below 70%. The following analyses represent data from the remaining 42 participants (25 female; *M* age = 18.76 years, *SD* = 0.98).

Materials and Procedure. Stimuli used in Experiment 2 were identical to those in Experiment 1. Study procedures were also identical except for a 3000 ms interval between the presentation of the final item and the recall phase. In this interval, participants verified whether a two-step math equation was solved correctly or incorrectly (based on the operation span task, Turner & Engle, 1989). Participants pressed the “Z” key for correctly solved equations and the “N” key for incorrectly solved equations. Keyboard keys were relabeled with “Y” and “N” respectively to counter any construct matching problems. This procedure mimics the one used by Atkins and Reuter-Lorenz (2008) in their investigations of false short-term memories.

Data Analysis. Data analysis procedures were the same as in Experiment 1.

Results

Mean accuracy on the math task was 0.70 (*SD* = 0.46). Participants who had lower than 70% accuracy on the math task were excluded from all analyses because low performance could have been due to ineffective distraction.

Participants correctly recalled an average of 7.41 words per trial (*SD* = 1.79). Serial position curves (see Figure 1B) reveal the expected pattern of enhanced recall for the first several positions as well as an attenuation of the increased recall previously seen at the last several positions. Again, the lowest recall occurred for the intermediate portions of the list.

Mean correct recall responses by sublist are plotted in Figure 2B. Responses varied reliably across sublist, *F*(2,82) = 20.30, *p* < .001. Participants accurately recalled more items from the "C" sublist than the "B" sublist, *t*(41) = -6.82, *p* <.001, and from the initial "A" sublist than the "B" sublist, *t*(41) = 4.82, *p* <.001. The number of correct recall responses between the "C" and "A" sublists were marginally significantly different from one another, *t*(41) = -1.87, *p* = .07.

Out of all errors, 34% were lure items, 24% were semantically related yet non-theme items, 1% were phonologic, and 9% were unrelated items. The remaining 32% of errors were repeated items (both within and between trials) and non-word utterances. Mispronunciations were not included either as errors or correct responses. Participants made an average of 0.03 (*SD* = 0.02) errors per trial. Mean recall errors by error type are plotted in Figure 3B. There was a significant main effect for error type (semantic, phonologic, unrelated) collapsing across sublist, *F*(2,82) = 108.11, *p* < .001. Participants made significantly more semantic errors than both phonologic, *t*(41) = 10.92, *p* <.001, and unrelated errors, $t(41) = 9.95$, $p <.001$. Additionally, they made more unrelated errors than phonologic errors, $t(41) = -7.05$, $p <.001$.

As in Experiment 1, there was a significant interaction between sublist (“A”, “B”, “C”) and error type (semantic, phonologic), $F(2,82) = 5.30$, $p = .007$. This interaction was driven by the differential incidence of semantic errors across sublists, $F(2,82) = 5.58$, $p =.005$, compared to the incidence of phonologic errors which were relatively rare and constant across sublists, $F(2,82) = 0.43$, $p = .65$. Differences between semantic and phonologic errors were significant at every sublist (ps < .001). Means for both semantic and phonologic errors are presented in Figure 4B. As can be seen in Figure 4B, semantic errors decreased significantly between the primacy and recency sublists, $t(41) = 3.01$, $p = .004$, whereas phonological errors remained relatively stable between the primacy and recency sublists, $t(41) = .221$, $p = .825$.

Discussion

Participants performed well on the recall task recalling nearly 60% of the 12 items presented. Although there was still a primacy effect at positions 1 and 2, the previously elevated performance at positions 11 and 12, as seen in Experiment 1, was noticeably diminished in Experiment 2. Therefore, the filled retention interval successfully impaired the recency effect.

Again, we observe a release from proactive interference effect (Underwood, 1945; Wickens, 1970) at the transition between the middle and recency sublists (positions 8 vs. 9) but not between the primacy and middle sublists (positions 4 vs. 5). We speculate that this is due to the filled retention interval because this pattern was not present in Experiment 1 where there was no retention interval or math distractor task.

Semantic errors were the most common type of errors made overall and in every sublist. The filled retention interval increased the amount of semantic errors from the “C” sublist while diminishing the amount of phonological errors relative to the baseline rates reported in the first experiment (see Experiment Comparison for further details). However, semantic errors from the primacy sublist still exceeded levels of semantic errors from the recency sublist.

This trade-off between decreased veridical performance coupled with increased semantic errors from the “C” sublist shows the interplay between verbatim and gist memory traces. Both traces are present; however, when both time and interfering information intercede, verbatim traces are less accessible and gist traces appear to predominate. The presence of semantic errors from the “C” sublist provide strong evidence that semantic codes are used in STM and therefore support unitary models that conceptualize similar stores for S/LTM.

Strict interpretation of the S/LTM distinction would interpret the interposed delay and interference introduced by the math verification task, as well as the similarities in performance between the “A” and “C” sublists, as evidence that all 12-items are being recalled from LTM. Yet, many working memory[1] tasks (e.g., operation span; Turner & Engle, 1989; Unsworth et al., 2005) include distraction before recall and yet still are interpreted as measuring working memory, or STM, performance. Additionally, Atkins and Reuter-Lorenz (2008) used both filled and unfilled retention intervals when testing STM and found reliable false memories under both conditions. Therefore, there is a precedent for interpreting responses following an interpolated delay and interference task

---

[1] In the present paper, we use the terms “working memory” and “short-term memory” interchangeably.

as within the province of STM. Direct comparisons between Experiments 1 and 2, and the implications for adjudicating between multiple- and unitary-store models, are discussed in the next section.

Experiment Comparison

Results

Participants recalled significantly fewer items per trial in Experiment 2 versus Experiment 1, $t(82) = 2.90$, $p = .005$. Additionally, the rate of correct recall responses was significantly lower in Experiment 2, $t(82) = 4.90$, $p < .001$.

When correct responses were broken down by sublist, there was a significant interaction between experiment (1, 2) and sublist (“A”, “B”, “C”), $F(1,82) = 24.51$, $p < .001$. This interaction was driven by significantly fewer items recalled from the “C” sublist in Experiment 2, $t(82) = 8.67$, $p < .001$, and a trend of fewer items recalled from the “B” sublist as well, $t(82) = 1.90$, $p = .06$ (see Figures 2A and 2B).

There was no significant main effect of experiment (1,2) on error type (semantic, phonologic, unrelated), $F(1,82) = 1.33$, $p = .25$. The 3-way interaction between experiment (1,2), error type (semantic, phonologic), and sublist (“A”, “B”, “C”) was not significant, $F(1,82) = 0.87$, $p = .36$. Based on *a priori* hypotheses, independent samples t-tests were conducted to examine differing rates of semantic and phonologic errors from the “C” sublist between the two experiments. There were significantly more semantic errors made from the “C” sublist following a filled retention interval, $t(82) = -4.07$, $p < .001$. While there were numerically less phonologic errors from the “C” sublist, this effect just reached significance, $t(82) = 1.99$, $p = .05$.

General Discussion

Both experiments showed that false memory errors, as indicated by the false recall of semantically related words, are associated with both primacy and recency portions of the serial position curve. In Experiment 1, as predicted by traditional multiple-store models, these errors were more frequent in the primacy sublist compared to the recency sublist whereas the opposite was true for phonological errors. Nevertheless, semantic errors constituted a sizable portion of the errors associated with the recency sublist which is strong evidence for the presence of semantic coding in association with STM. This interpretation is strengthened by the finding that semantic influences were accentuated by introducing a filled retention interval in Experiment 2.

The view that false memories are due to associative activation assumes that these semantic errors result from semantic processing at encoding. Specifically, during encoding, list items activate a network of semantic relatives (Meyer & Schvaneveldt, 1971; Roediger et al., 2001; Underwood, 1965). This view is consistent with our finding that participants were most likely to recall the highest normative semantic associate, the non-presented theme word, or a word judged to be semantically related to the studied list. Observing these semantic errors in both the primacy and recency sublists argue for the operation of semantic encoding at multiple points in the 12-item list.

Another way to interpret these findings is in terms of the relationship between accurate memory performance and false memory errors. The filled retention interval significantly changed the relative predominance of semantic errors and veridical recall from the “C” sublist. This shows that both verbatim and gist memory traces are present, but vary in strength, based on the other demands of the task. These data suggest a trade-off between veridical memory performance and semantic false memories. Veridical memories are supported by verbatim and gist traces whereas false memories are supported by gist traces in the absence of verbatim traces (Brainerd & Reyna, 2002). The occurrence of both veridical and semantic false memories from within the “C” sublist, as well as at other sublist positions, indicates that there must be both verbatim and gist traces present. This outcome is inconsistent with a strong multiple-store view that presumes minimal semantic coding in STM.

However, we acknowledge that phonologic errors were more prevalent from the “C” sublist than any other sublist. Shallow, perceptually-based codes are available in the recency sublist of the curve more so than from the primacy sublist. But, since this variable was not directly manipulated, no conclusions can be drawn from the present study. We take the results of these experiments as evidence that S/LTM are overlapping, but not identical, systems in memory.

Future Directions

This study shows strong behavioral evidence for a common underlying code between S/LTM. While behavioral observations can provide a window into what the mind may be doing, they cannot always capture the complexities of the brain processes underlying these mental processes. Functional neuroimaging methods, such as PET (positron emission tomography) and fMRI (functional magnetic resonance imaging), allow another method of peering into the “black box” of the mind and often can generate insights that pure behavioral work cannot (e.g., Jonides, Nee, & Berman, 2006). Also, at different points, neuroimaging has been used as evidence for separate stores (e.g., Talmi, et at., 2005) and as evidence against these theories in favor of unitary stores (Ranganath & Blumenfeld, 2005). A limitation of the present study would be that this exact task is not appropriate for the scanning environment given that vocal responses, as required by a recall task, induce head motion and therefore noise into the imaging signal. Development of a recognition version of this task would be useful for direct comparison to Talmi et al. (2005) with an fMRI task.

A recognition task would also allow for the direct investigation of the recency sublist immediately following list presentation. In free recall, responses are unconstrained therefore reporting items from other portions of the list, prior to the recency items, may interfere with overall recall of the recency sublist. In a recognition version of the task, we would predict high veridical memory performance from all sublists given that recognition tasks generally yield heightened accuracy in comparison to recall tasks due to easier response generation demands (Sternberg, 2006, p.158). Furthermore, we would predict that semantic errors will be present from all sublists but at the greatest levels from the primacy sublist—similar to our findings in Experiment 1. Verbatim traces should be strongest from the recency sublist and these should override the gist traces thus leading to high veridical memory coupled with low false memories. Recent evidence from Atkins and Reuter-Lorenz (2008) and Flegal et al. (in press) provide evidence that, even in immediate recognition, STM is vulnerable to semantic distortions.

The connections between S/LTM are important for understanding the structure of memory, but, another important concern is how memory changes across the lifespan. The present study is limited in that it can only speak to the relation of S/LTM in a healthy, younger adult population. Given the importance of memory and the insults it faces with age (e.g., Alzheimer’s disease; Nebes, 1992), understanding changes in memory across the lifespan may be important for better understanding of how to treat age-related memory declines.

Another way that false memory research can help improve real-world memory is by identifying other cognitive domains that relate to false memories, for example working memory capacity (e.g., Watson, Bunting, Poole, & Conway, 2005), that may be trainable processes that can be leveraged to improve and rehabilitate memory.

Conclusions

This goal of this research was to answer the question of whether unitary or separable systems mediate the ability to remember new numbers and our own phone numbers over extended durations. To answer this question, we looked at the rates and types of memory failures in an adaptation of a classic list learning task. The data we have presented here demonstrate the occurrence of semantic memory errors across sublists, which suggest the operation of common, shared operations in S/LTM. Specifically, we take the presence of semantic errors in a traditionally STM domain (the recency portion of the serial position curve) as an indication that the same deep, meaning-based code underlying LTM is operating in STM as well. Future experimentation using recognition in conjunction with neuroimaging can be used to corroborate these conclusions.

References

Alpert, J.L., Brown, L.S., Ceci, S.J., Courtois, C.A., Loftus, E.F., & Ornstein, P.A. (1998). Final conclusions of the American Psychological Association working group on investigation of memories of childhood abuse. *Psychology, Public Policy, and Law, 4*, 933-940.

Anastasi, J. S., Rhodes, M. G., & Burns, M. C. (2000). Distinguishing between memory illusions and actual memories using phenomenological measurements and explicit warnings. *American Journal of Psychology, 113*(1), 1-26.

Atkinson, R. C. & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In Spence, K. W. & Spence, J. T. (Eds.), *The Psychology of Learning and Motivation: Advances in Research and Theory, Volume 2*. (pp. 89-195). New York: Academic Press.

Atkins, A.S. & Reuter-Lorenz, P.A. (2008). False working memories? Semantic distortion in a mere 4 seconds. *Memory and Cognition, 36*(1), 74-81.

Baddeley, A.D. (1966). The influence of acoustic and semantic similarity on long-term memory for word sequences. *The Quarterly Journal of Experimental Psychology, 18*, 302-309.

Baddeley, A. D. (2003). *Essentials of Human Memory* (p.32). New York: Psychology Press Limited.

Baddeley, A.D. & Warrington, E. K. (1970). Amnesia and the distinction between long- and short-term memory. *Journal of Verbal Learning and Verbal Behavior, 9*, 176-189.

Bahrick, H.P. (1984). Fifty years of second language attrition: Implications for programmatic research. *The Modern Language Journal, 68*, 105-118.

Brainerd, C.J. & Reyna, V.F. (2002). Fuzzy-trace theory and false memory. *Current Directions in Psychological Science*, *11*, 164-169.

Cabeza, R., Dolcos, F., Graham, R., & Nyberg, L. (2002). Similarities and differences in the neural correlates of episodic memory retrieval and working memory. *Neuroimage*, *16*, 317-330.

Cowan, N. (2000). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*, 87-185.

Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology, 58*, 17-22.

Deese, J., & Kaufman, R.A. (1957). Serial effects in recall of unorganized and sequentially organized verbal material. *Journal of Experimental Psychology*, *54*, 180-187.

Drag, L.D., Bieliauskas, L.A., Kaszniak, A.W., Bohnen, N.I., & Glisky, E.L. (2009). Source memory and frontal functioning in Parkinson’s disease. *Journal of International Neuropsychological Society, 15*, 399-406.

Ebbinghaus, H. (1885). *On memory* (H.A. Ruger & C.E. Bussenius, Trans.). New York: Teachers’ College, 1913.

Flegal, K.E., Atkins, A.S. & Reuter-Lorenz, P.A. (in press). False memories seconds later: The rapid and compelling onset of illusory recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition.*

Gallo, D.A. (2006). *Associative illusions of memory: False memory research in DRM and related tasks.* New York: Psychology Press.

Gallo, D.A., Roediger, H.L. III, & McDermott, K.B. (2001). Associative false recognition occurs without strategic criterion shifts. *Psychonomic Bulletin & Review, 8,* 579-586.

Glanzer, M. & Cunitz, A.R. (1966). Two storage mechanisms in free recall. *Journal of Verbal Learning and Verbal Behavior, 5,* 351-360.

Hannula, D.E., Tranel, D., & Cohen, N.J. (2006). The long and the short of it: Relational memory impairments in amnesia, even at short lags. *The Journal of Neuroscience, 26*, 8352-8359.

Heaps, C.M., & Nash, M. (2001). Comparing recollective experience in true and false autobiographical memories. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*, 920-930.

Hyman, I.E., Husband, T.H., & Billings, F.J. (1995). False memories of childhood experiences. *Applied Cognitive Psychology, 9,* 181-197.

Jonides, J., Lewis, R.L., Nee, D.E., Lustig, C.A., Berman, M.G., & Moore, K.S. (2008). The mind and brain of short-term memory. *Annual Review of Psychology, 59*(1), 193-224.

Jonides, J., Nee, D.E., & Berman, M.G., (2006). What has functional neuroimaging told us about the mind? So many examples, so little space. *Cortex, 42,* 414-417.

Laming, D. (2010). Serial position curves in free recall. *Psychological Review, 117*(1), 93-133.

Loftus, E.F. (1997). Creating false memories. *Scientific American, 277*(3), 70-78.

Loftus, E.F. & Pickrell, J.E., (1995). The formation of false memories. *Psychiatric Annals, 25*, 720-725.

Melton, A.W. (1963). Implications of short-term memory for a general theory of memory. *Journal of Verbal Learning and Verbal Behavior, 2*(1), 1-21.

Meyer, D.E. & Schvaneveldt, R.W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology, 90,* 227-234.

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

Milner, B. (1978). Clues to the cerebral organization of memory. In Buser, P. A. & Rouged-Buser, A. (Eds.), *Cerebral Correlates of Conscious Experience* (pp.139-153). Elsevier: Amsterdam.

Murdock, B.B. (1962). The serial position effect on free recall. *Journal of Experimental Psychology, 64,* 482-488.

Nebes, R. D. (1992). Cognitive dysfunction in Alzheimer’s Disease. In F.I.M. Craik & T.A. Salthouse (Eds.), *The handbook of aging and cognition* (pp. 373-446). New Jersey: Hove and London.

Payne, D. G., Elie, C. J., Blackwell, J. M., & Neuschatz, J. S. (1996). Memory illusions: Recalling, recognizing, and recollecting events that never occurred. *Journal Of Memory and Language, 35,* 261-285.

Postman, L. & Phillips, L.W. (1965). Short-term temporal changes in free recall. *Quarterly Journal of Experimental Psychology, 17*, 132-138.

Ranganath, C. & Blumenfeld, R. S. (2005). Doubts about double dissociations between short- and long-term memory. *Trends in Cognitive Sciences, 9*, 374-380.

Roediger, H.L.III & McDermott, K.B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology, 21*, 803-814.

Roediger, H.L.III, Watson, J.M., McDermott, K.B., & Gallo, D.A. (2001). Factors that determine false recall: A multiple regression analysis. *Psychonomic Bulletin and Review, 8*, 385-407.

Schacter, D.L. (1999). The seven sins of memory: Insights from psychology and cognitive neuroscience. *American Psychologist, 54*, 182-203.

Scoville, W.B. & Milner, B. (1957) Loss of recent memory after bilateral hippocampal lesions. *Journal of Neurological Neurosurgery Psychiatry, 20*, 11–21.

Seamon, J.G., Luo, C.R., Kopecky, J.J., Price, C.A., Rothschild, L., Fung, N.S., & Schwartz, M.A. (2002). Are false memories more difficult to forget than accurate memories? The effect of retention interval on recall and recognition. *Memory and Cognition, 30*, 1054-1064.

Shallice, T. & Warrington, E.K. (1970). Independent functioning of verbal memory stores: A neuropsychological study. *Quarterly Journal of Experimental Psychology, 22*, 261-273.

Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs: General and Applied, 74*, 1-28.

Sternberg, R.J. (2006). *Cognitive Psychology, 4th ed.* (p.158). Belmont: Thomson Wadsworth.

Talmi, D., Grady, C. L., Goshen-Gottstein, Y., & Moscovitch, M. (2005). Neuroimaging the serial position curve: A test of single-store versus dual-store models. *Psychological Science, 16*, 716-723.

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

Underwood, B.J. (1945). The effect of successive interpolations on retroactive and proactive inhibition. *Psychological Monographs, 59*(3).

Underwood, B.J. (1965). False recognition produced by implicit verbal responses. *Journal of Experimental Psychology, 70*(1), 122-129.

Unsworth, N., Heitz, R.P., Schrock, J.C., & Engle, R.W. (2005). An automated version of the operation span task. *Behavior Research Methods, 37*, 498-505.

Vallar, G. & Shallice, T. (Eds.) (1990). *Neuropsychological Impairments of Short-Term Memory* (pp.11-53). Cambridge: Cambridge University Press.

Watson, J.M., Bunting, M.F., Poole, B.J., & Conway, A.R.A. (2005). Individual difference in susceptibility to false memory in the Deese-Roediger-McDermott paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*(1), 76-85.

Wechsler, D. (1997). *WAIS-III administration and scoring manual*. The Psychological Corporation, San Antonio, TX.

Wickelgren, W.A. (1965). Memory for phonemically similar lists. *American Journal of Psychology, 78,* 567-574.

Wickens, D.D. (1970). Encoding categories of words: An empirical approach to meaning. *Psychological Review, 77,* 1-15.

Author Note

Halle R. Zucker, Department of Psychology. University of Michigan, Ann Arbor.

This research was funded by a LS&A Honors Grant, Psychology Department Conference Travel Award, and the Tanner Memorial Award. These funds were used to pay subjects, purchase materials, and support conference presentations. Data from this project will be presented at the 2010 meeting of the Cognitive Neuroscience Society.

I would like to thank Tom Bergman, Tamara Brodsky, Ashley Brooks, Jackie Dobson, Autumn Gear, Rachel German, Katherine Khatibi, Michael Kinning, Nina Massad, and Julia Tattan without whose help I would still be transcribing and coding audio files. I also wish to thank Alex Atkins and Kristin Flegal for their emotional and professional support over the past four years and for doing such great science before me with this paradigm—I couldn't ask for better role models. Thank you to Sara Festini and Lynn Ossher for their help with statistical analyses.

I also owe a huge debt to my friends and family who supported me throughout this process and who often bore the brunt of my stresses and frustrations.

Thank you to Patti Reuter-Lorenz for being a wonderful mentor and role model over the past four years and especially throughout the process of writing this thesis. Thank you to the other members of the Reuter-Lorenz lab, Kristi MacKenzie in the honors office, and all of the other people who made this research possible.

Correspondence concerning this manuscript should be addressed to H. R. Zucker, Department of Psychology, University of Michigan, 530 Church St., Ann Arbor, MI 48109 or via e-mail to hallez@umich.edu.

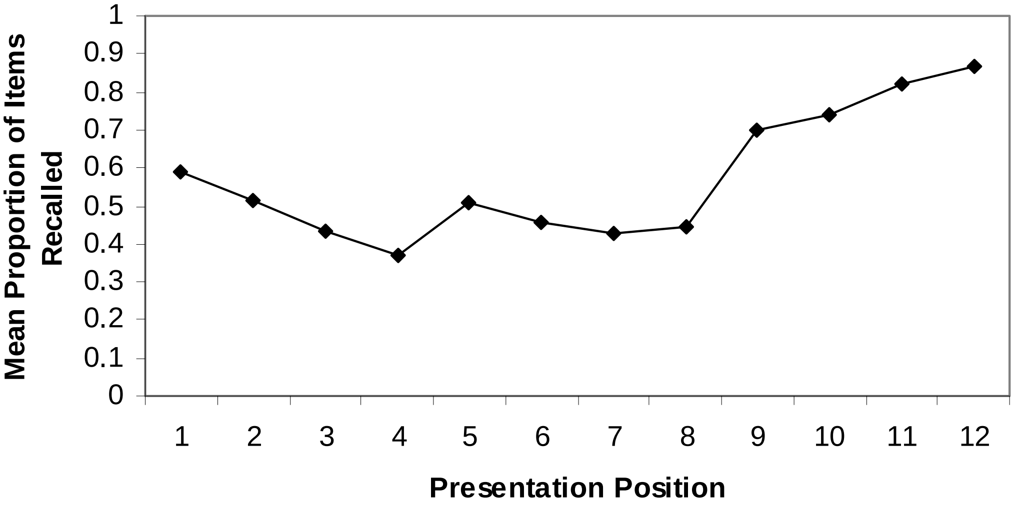
Table 1

Digit Span, Operation Span, and Source Memory Scores

Task	Experiment 1 Mean (*SD*)	Experiment 2 Mean (*SD*)
Digit Span Forwards	11.29 (2.05)	11.33 (2.18)
Digit Span Backwards	7.95 (2.14)	7.57 (2.45)
Digit Span Total	19.24 (3.68)	18.90 (3.91)
Operation Span	60.36 (10.54)	61.34 (8.73)
Source Memory	13.64 (2.23)	13.85 (2.69)

Note. Digit span total represents the sum of digit span forwards (max score = 16 points) and digit span backwards (max score = 14 points). Operation span scores were taken out of 80 points (after Unsworth et al., 2005). Source memory scores were taken out of 20 points.

a.



b.

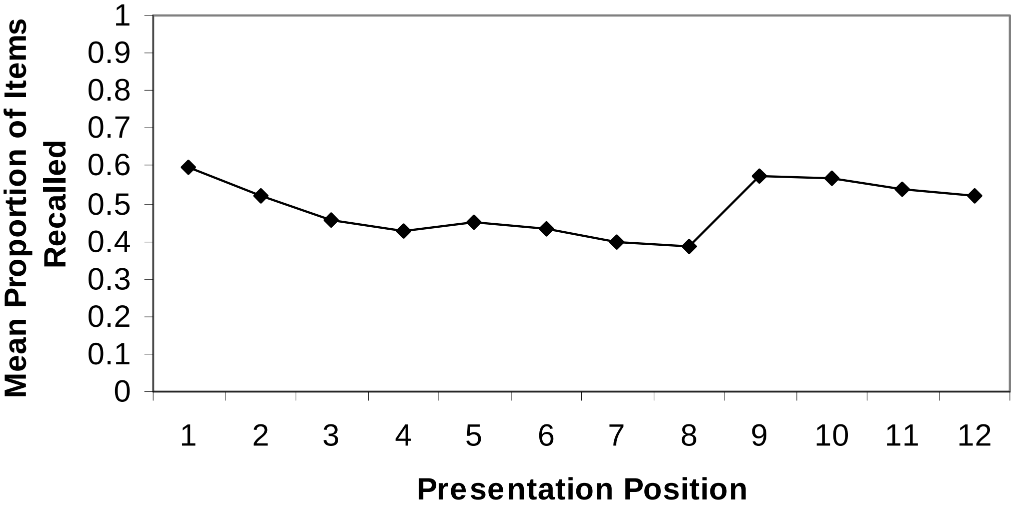
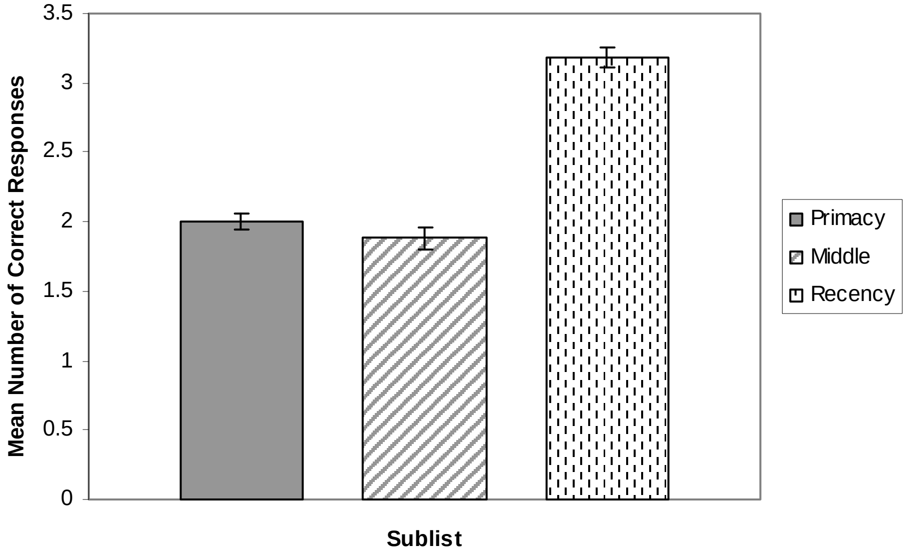


Figure 1. Mean proportion of items recalled from each presentation position. (a) The results from Experiment 1 show the classic primacy and recency effects associated with positions 1 and 2, and 11 and 12 respectively. In addition, a release from proactive interference effect appears to be superimposed as evident in recall enhancement in the transitions between sublists—positions 4 vs. 5 and 8 vs. 9. (b) The inclusion of a filled retention interval in Experiment 2 did not alter the primacy effect associated with positions 1 and 2 but diminished the recency effect at positions 11 and 12. In addition, a release from proactive interference effect occurs between positions 8 and 9 but not between positions 4 and 5. We speculate this absence between positions 4 and 5 was due to the filled retention interval.

a.



b.

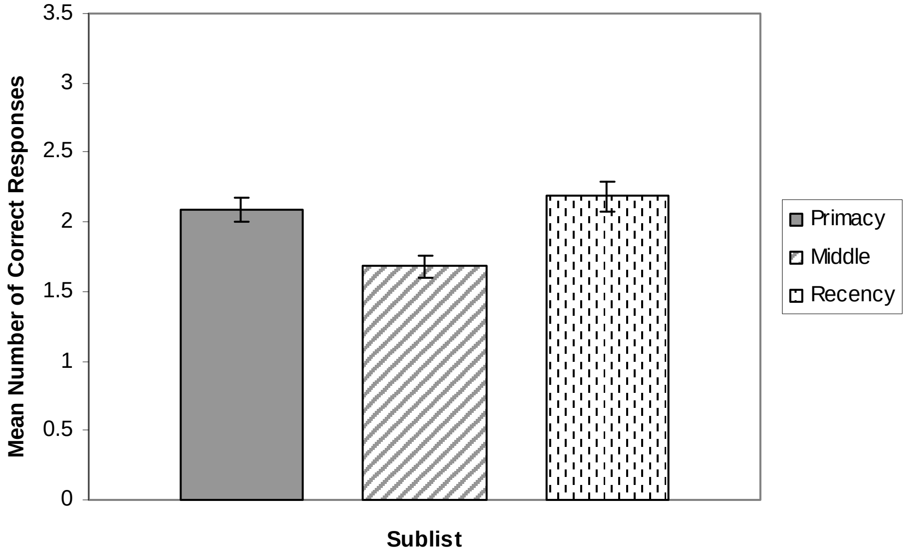
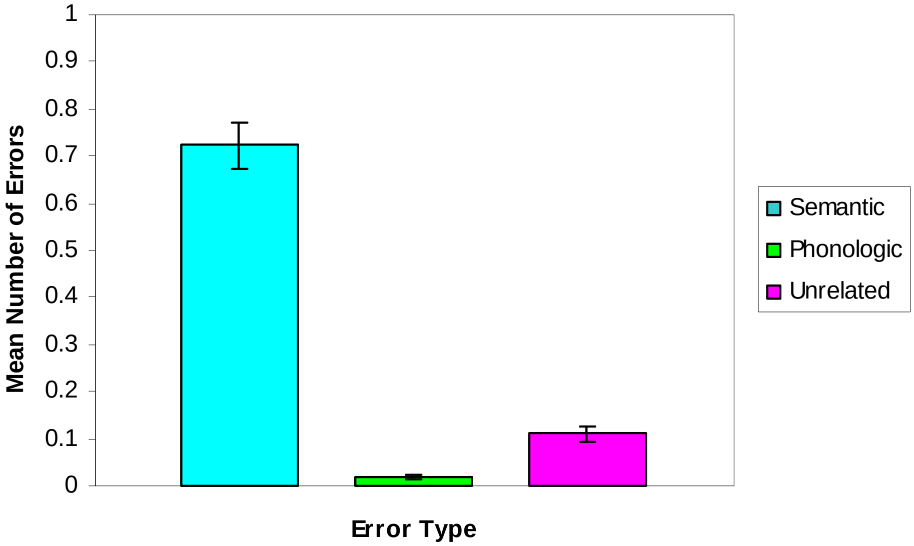


Figure 2. Mean correct responses by sublist (error bars = *SEM*). For both experiments, there was a main effect of sublist, *p*s < .001. (a) The results from Experiment 1 show that responses from the recency sublist (“C”) varied reliably from both the primacy (“A”) and middle (“B”) sublists, *p*s <.001. Responses from the primacy and middle sublists did not reliably differ, *p* = .11. (b) The inclusion of a filled retention interval in Experiment 2 caused responses to vary reliably between the primacy and middle, and recency and middle sublists, *p*s <.001, but only marginally between the primacy and recency sublists, *p* = .07.

a.



b.

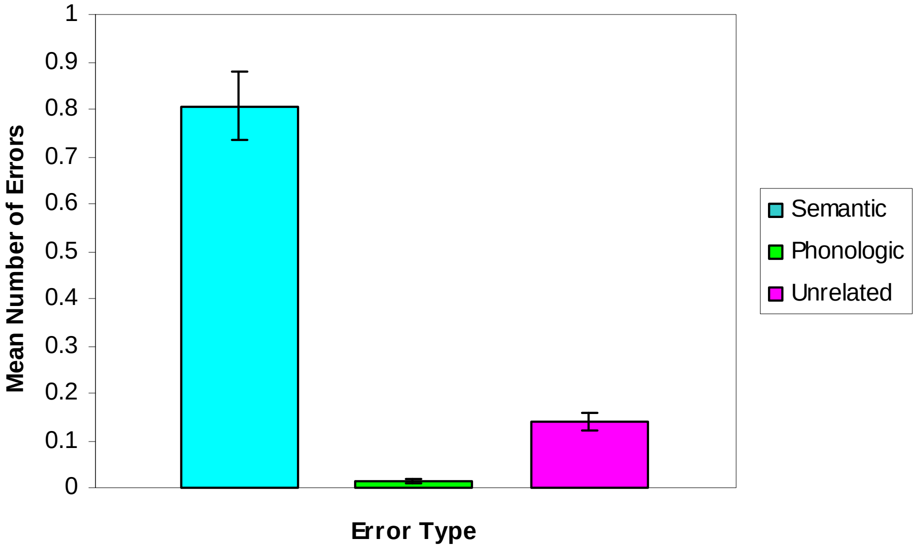
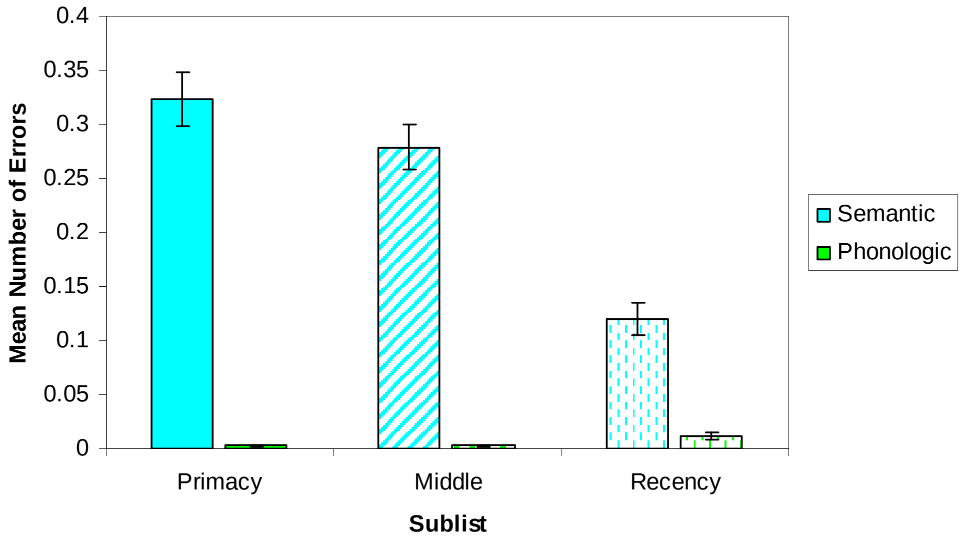


Figure 3. Mean recall errors by error type (semantic, phonologic, unrelated) from Experiment 1 (a) and 2 (b). In both experiments, there was a significant main effect of error type, *p*s < .001. In each experiment, all error types were significantly different from every other error type, *p*s < .001. (Error bars = *SEM*).

a.



b.

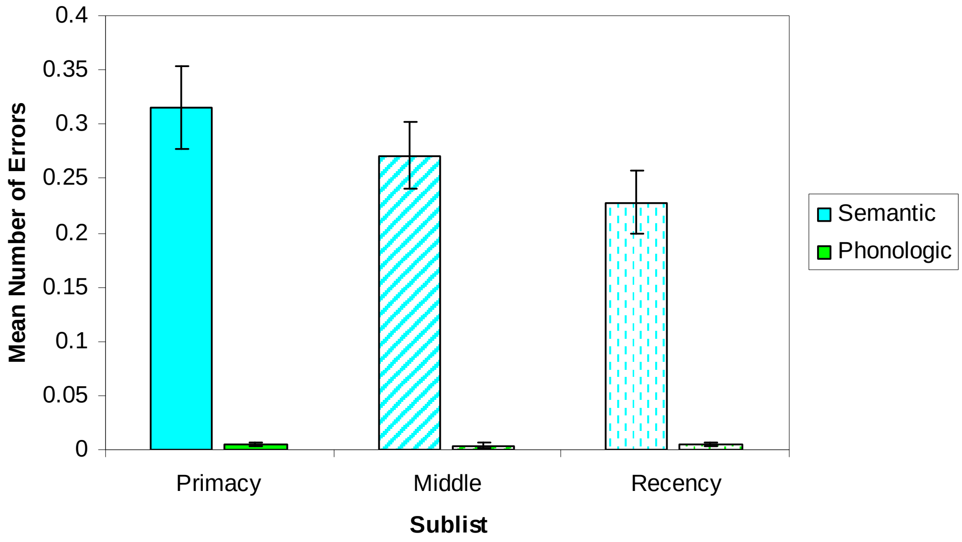


Figure 4. Mean recall responses by error type (semantic, phonologic) across sublists (“A”, “B”, “C”; error bars = *SEM*). In both experiments, there was a significant interaction between sublist and error type, *p*s <.001, which was significant at every sublist, *p*s <.001. (a) In Experiment 1, there was a main effect of sublist on both semantic, *p* <.001, and phonologic errors, *p* = .001. (b) With inclusion of a filled retention interval in Experiment 2, the main effect of sublist on semantic errors, *p* = .005, persisted but was abolished for phonologic errors, *p* = .65.

Appendix

The Forty-Two 12-item Lists Used in Experiments 1 & 2, Listed Alphabetically by Triplet

ACT-SHOVEL-SLEEP
perform
portray
drama
pretend
dig
spade
pail
rake
nap
doze
bed
awake

ADJECTIVE-SNAKE-SOFT
adverb
pronoun
noun
verb
slither
serpent
reptile
rattle
hard
loud
tender
fluffy

ADULT-BLOOD-SHIRT
mature
responsible
grown
kid
plasma
donor
vein
vampire
blouse
sleeves
collar
shorts

AGAIN-BARBECUE-FRIEND
repeat
never
twice
stop
cookout
grill
ribs
sauce
pal
buddy
companion
neighbor

ALONE-ARMY-FOG
isolated
solo
lonely
one
navy
soldier
infantry
marines
mist
haze
smog
unclear

ANGER-SNEEZE-STOMACH
rage
mad
enrage
fury
allergy
cough
handkerchief
tissue
abdomen
belly
intestine
ulcer

ANNOY-CITY-ROCK
bother
aggravate
irritate
disturb
metropolis
town
urban
suburb
boulder
stone
solid
roll

ANSWER-BLACK-CABBAGE
question
reply
response
solution
white
gray
brown
coal
patch
sauerkraut
slaw
lettuce

ARGUE-SINK-SPIDER
debate
disagree
complain
agree
drain
float
faucet
bathroom
web
tarantula
arachnid
creepy

ATOM-BROOM-LIE
molecule
nucleus
neutron
proton
dustpan
sweep
mop
witch
fib
untruthful
deception
dishonest

AUTHOR-LOSE-NEEDLE
writer
poet
editor
publisher
win
find
gain
defeat
thread
syringe
haystack
injection

BABY-BASKET-THIEF
crib
infant
diaper
carriage
wicker
picnic
waste
laundry
crook
robber
burglar
bandit

BACK-GAS-HELP
front
spine
behind
forward
fuel
petroleum
station
oil
assist
aid
emergency
wanted

BAKE-KING-ROOF
broil
oven
cook
cake
throne
queen
crown
reign
shingle
ceiling
tar
tin

BEAUTIFUL-CORN-SLOW
gorgeous
lovely
pretty
ugly
cob
husk
flake
field
fast
snail
turtle
sluggish

BEE-BUY-WISH
hive
bumble
sting
buzz
purchase
sell
store
spend
hope
want
desire
dream

BOOM-FISH-TEETH
sonic
bang
explosion
bomb
trout
cod
scales
shrimp
gums
braces
mouth
tongue

BOX-CHURCH-TRIP
cardboard
carton
storage
container
cathedral
steeple
temple
preacher
journey
vacation
travel
baggage

BREAD-MOUNTAIN-RENT
rye
loaf
butter
toast
climber
hill
climb
molehill
own
lease
apartment
monthly

BUILDING-MORNING-SMELL
structure
blocks
construction
empire
early
dawn
dew
evening
aroma
scent
whiff
stench

BUTTERFLY-COLD-PAN
cocoon
moth
insect
wing
hot
shiver
arctic
frigid
skillet
pot
fry
dish

CARPET-MAN-WINDOW
rug
floor
magic
red
woman
lady
handsome
male
pane
sill
shutter
curtain

CHAIN-FAIL-JOB
link
whip
necklace
bicycle
flunk
pass
succeed
try
occupation
employment
career
task

CHAIR-PARK-ROPE
table
rocking
swivel
recliner
lot
bench
recreation
playground
knot
string
noose
twine

CHAOS-FOOT-FRUIT
havoc
anarchy
hectic
confusion
toe
inch
ankle
shoe
kiwi
citrus
pear
berry

CHEESE-LION-MUSCLE
cheddar
swiss
cracker
mouse
roar
tamer
tiger
mane
flex
weights
strength
tone

CLAM-DESTROY-FINISH
chowder
oyster
shell
mussel
demolish
ruin
annihilate
create
done
start
complete
end

COPY-EGGS-FRAGILE
duplicate
carbon
original
photo
omelet
bacon
dozen
scramble
delicate
breakable
frail
glass

CUP-PEN-PULL
saucer
measuring
mug
goblet
quill
pencil
marker
write
tug
push
drag
stretch

DANCE-MARRY-MOVIE
ballet
ballerina
song
aerobics
wed
engage
single
hitch
cinema
film
theater
popcorn

DOCTOR-DRY-GIRL
physician
nurse
stethoscope
surgeon
towel
desert
moist
thirst
boy
dolls
female
dress

FLAG-RIVER-SMOKE
banner
checkered
stripes
pole
creek
stream
flow
bridge
cigar
cigarette
pipe
tobacco

FLOWER-MAP-WHOLE
tulip
petals
daisy
vase
atlas
chart
direction
world
half
part
piece
all

FOREVER-GHOST-JUSTICE
eternity
infinity
always
endless
ghoul
goblin
phantom
spook
liberty
courts
truth
lawyer

FUNNY-HEALTH-PIG
hilarious
comedian
humor
clown
sickness
body
wealth
ill
hog
pork
sow
sty

GIVE-HIGH-NUT
take
generous
share
charity
low
elevate
tower
jump
cashew
pecan
almond
squirrel

HOLD-PRESENT-SQUARE
grasp
grip
keep
carry
gift
past
absent
future
circle
triangle
round
shape

HORSE-PIANO-RUBBER
saddle
pony
gallop
colt
keyboard
organ
guitar
ivory
foam
latex
galoshes
tire

LETTER-RING-TRASH
envelope
stamp
mailbox
mail
diamond
bell
jewelry
phone
garbage
rubbish
debris
dump

MANY-RAIN-ROUGH
several
few
much
plenty
umbrella
storm
hail
puddle
sandpaper
smooth
coarse
tough

MATH-SAFE-STRESS
arithmetic
calculus
algebra
equation
vault
secure
guard
lock
tension
pressure
anxiety
strain

SMART-SPICE-VOTE
intelligent
genius
wise
knowledge
oregano
herb
cinnamon
seasoning
ballot
election
register
campaign