The Effect of Concurrent Semantic Categorization on Delayed Serial Recall

Daniel J. Acheson, Maryellen C. MacDonald, and Bradley R. Postle
University of Wisconsin—Madison

The influence of semantic processing on the serial ordering of items in short-term memory was explored using a novel dual-task paradigm. Participants engaged in 2 picture-judgment tasks while simultaneously performing delayed serial recall. List material varied in the presence of phonological overlap (Experiments 1 and 2) and in semantic content (concrete words in Experiment 1 and 3; nonwords in Experiments 2 and 3). Picture judgments varied in the extent to which they required accessing visual semantic information (i.e., semantic categorization and line orientation judgments). Results showed that, relative to line-orientation judgments, engaging in semantic categorization judgments increased the proportion of item-ordering errors for concrete lists but did not affect error proportions for nonword lists. Furthermore, although more ordering errors were observed for phonologically similar relative to dissimilar lists, no interactions were observed between the phonological overlap and picture-judgment task manipulations. These results demonstrate that lexical-semantic representations can affect the serial ordering of items in short-term memory. Furthermore, the dual-task paradigm provides a new method for examining when and how semantic representations affect memory performance.

Keywords: working memory, language production, serial recall, concreteness, speech error

Working memory (WM) refers to the ability to maintain and manipulate information over short periods of time. For verbal information, maintenance (i.e., short-term memory; STM) typically is thought to occur over a strictly phonological level of representation within specialized, short-term memory stores whose contents either decay or are interfered with over time (e.g., the phonological loop; Baddeley, 1986). This view is based largely on findings showing relatively large phonological and weaker semantic effects on verbal STM performance (e.g., Baddeley, 1966), as well as on patient populations who exhibit severe deficits in STM while exhibiting relatively spared language comprehension and production (Shallice & Warrington, 1970; Vallar & Baddeley, 1984). However, this conception of verbal STM may be skewed by virtue of the fact that the vast majority of research has used stimuli (e.g., letters, digits; Conrad & Hull, 1964; Wickelgren, 1965) or testing conditions (randomly sampling from closed word lists; e.g., Baddeley, 1966) that likely minimize the use of semantic processing. It is critical that when lexical-semantic factors have been systematically investigated, they have been shown to influence STM performance. For instance, individuals show superior recall for words over nonwords (Brener, 1940; Hulme, Maughan, & Brown, 1991), for frequent relative to infrequent words (Gregg, Freedman, & Smith, 1989; Hulme et al., 1997; Watkins, 1977), and for concrete/imageable words relative to abstract/nonimageable ones (Bourassa & Besner, 1994; Walker & Hulme, 1999). Thus, both phonological and lexical-semantic factors affect verbal STM, yet relatively few studies have manipulated both factors simultaneously. The present investigation is designed to address the influence of lexical-semantic and phonological representations on the maintenance of information in verbal WM through the use of a novel dual-task paradigm.

One of the reasons that phonological and lexical-semantic influences on WM performance have been viewed independently is that they appear to have different effects on order and item memory, respectively. Manipulations of phonological overlap (i.e., phonological similarity), for instance, affect people's memory for the order in which items appear but not memory for the items themselves (Fallon, Groves, & Tehan, 1999). On the other hand, semantic similarity among list items improves memory for semantically grouped relative to ungrouped items with no apparent influence on memory for the order in which they appeared (Poirier & Saint-Aubin, 1999; Saint-Aubin & Poirier, 2000; cf. Baddeley, 1966). Written frequency shows a comparable result, with superior item memory for high- relative to low-frequency words (Hulme et al., 1997). Moreover, neuropsychological investigations have re-

vealed individual patients with selective disruption of phonological and semantic processing, respectively (Martin, Lesch, & Bartha, 1999; Martin, Shelton, & Yaffee, 1994). Findings such as these have produced a standard assumption in verbal WM research that phonological factors influence short-term, serial ordering processes, whereas lexical-semantic representations influence item memory via contributions from long-term memory (Baddeley, 1972; cf. Crowder, 1989).

Two very different theoretical perspectives have arisen to explain how lexical-semantic and phonological codes influence verbal WM. *Trace redintegration* accounts retain the classic distinction between short- and long-term memory (e.g., Atkinson & Shiffrin, 1968). According to this and related accounts, to-be-remembered information is maintained in a purely phonological form (e.g., the phonological loop; Baddeley, 1986), and it is only during the time of recall that lexical-semantic representations stored in long-term memory are used to clean up, or redintegrate a degraded phonological representation (Hulme et al., 1997; Schweickert, 1993). In most theories that use such a mechanism, redintegration is thought to occur on an item-by-item basis, which explains why lexical-semantic information has its influence on item and not order recall (cf. Stuart & Hulme, 2000). An alternative is provided by accounts in which maintenance is achieved via temporary activation of long-term memory (e.g., Cowan, 1995; Crowder, 1993; Oberauer, 2002; Postle, 2006; Ruchkin, Grafman, Cameron, & Berndt, 2003). In the case of verbal WM specifically, maintenance occurs over not just phonological but also lexical-semantic and potentially other levels of linguistic representation (Acheson & MacDonald, 2009b; Haarmann & Usher, 2001; Martin, et al., 1999; Saffran & Martin, 1997). Although there is some disagreement regarding the extent to which maintenance requires specialized storage buffers (Martin, et al., 1999) or not (Acheson & MacDonald, 2009b; Haarmann & Usher, 2001; Saffran & Martin, 1997), these perspectives generally hold that lexical-semantic representation is maintained along with phonological representation. Furthermore, some suggest that the processing occurring over lexical and semantic representations can influence not just memory for the items themselves but also memory for serial order (e.g., Acheson & MacDonald, 2009b; Saffran & Martin, 1997).

One such account in which semantic representations play a critical role in the serial ordering of phonological representations is the *semantic binding hypothesis* (Patterson, Graham, & Hodges, 1994). This hypothesis states that two sources of information bind the phonological elements of words together: the learned, co-occurrence of speech sounds within a word and the association of a word with its respective semantics. Evidence for this account comes from studies of patient populations who suffer from a progressive loss of semantic representation and processing (i.e., semantic dementia patients; Hodges, Patterson, Oxbury, & Funnell, 1992). Within such patient groups, memory for list material that is known to the individual is superior to memory for unknown material (Knott, Patterson, & Hodges, 2000; Patterson et al., 1994). A particularly striking feature of patient performance is that the recall of the unknown words often results in numerous phonological errors in which the patients mix together speech sounds from the to-be-remembered material (e.g., recalling *mint, rug* as *r*int, *m*ug). Similar patterns of performance have been observed in patient populations with semantic impairments stemming from different sources (Caza, Belleville, & Gilbert, 2002; Forde &

Humphreys, 2002; Wong & Law, 2008). Although it is often difficult to induce comparable, sublexical serial ordering errors in nonpatient populations, such errors have been observed in situations when to-be-remembered stimuli are either devoid of semantic representation (e.g., nonwords; Acheson & MacDonald, 2009a; Treiman & Danis, 1988) or when there is a low ratio of words to nonwords in a to-be-remembered list (Jefferies, Frankish, & Lambon Ralph, 2006b).

In addition to effects of semantics on sublexical serial ordering processes, there is also evidence that lexical-semantic representation may influence the serial ordering of whole items. For instance, in normal language production, word exchanges in a sentence (e.g., I wrote a *mother* to my *letter*) almost always occur between words of the same grammatical category (Dell & Reich, 1981), suggesting that shared semantic representation may influence the relative activation of lexical representations and, hence, the order in which they are produced. In the context of STM task performance, individuals show an advantage for concrete over abstract word lists in an order-reconstruction task (Romani, McAlpine, & Martin, 2008). In addition, Acheson, Postle, and MacDonald (2010) demonstrated an intact phonological similarity effect for visually presented concrete words under conditions of concurrent articulation. Previous research has shown that the phonological similarity effect is abolished when stimuli are presented visually under conditions of concurrent articulation (Levy, 1971; Murray, 1968); Acheson et al. showed that the semantic properties of the stimuli matter. Although the phonological similarity effect was abolished for visually presented, abstract word lists, it remained for concrete word lists and was driven specifically by item-ordering errors. These results thus demonstrate that the serial ordering of items in verbal WM is determined not just by phonological factors but by lexical-semantic ones as well.

The semantic binding hypothesis offers a clear explanation of how lexical-semantic information can affect sublexical serial ordering. What remains unclear is how such representation might influence the serial ordering of whole items. Recent hypotheses about relationships between language production and verbal WM maintenance offer some insight into this mechanism (e.g., Acheson & MacDonald, 2009b; Allen & Hulme, 2006; Page, Madge, Cumming, & Norris, 2007). For instance, Acheson and MacDonald (2009b) suggested that maintenance verbal WM might arise via temporary activation of representations within the language production architecture, and Page et al. (2007) likened serial recall to a speech *re*production task, requiring maintenance over a lexical-level utterance plan. Performing a memory task is not exactly like typical language production, because in recall and recognition tasks, there is no coherent message which drives the generation of an ordered sequence of lexical items, as the order of the items is determined by the experimenter, not by the speaker/participant. As Page et al. (2007) noted, however, there is a lexical-level utterance plan. Critically, within essentially all models of language production, such an utterance plan either includes semantic representation exclusively (i.e., lemmas; Levelt, Roelofs, & Meyer, 1999) or some combination of the semantic and phonological influences (e.g., interactive activation accounts; Dell, 1986; Dell & O'Seaghdha, 1992; Foygel & Dell, 2000). These interactive activation frameworks provide a potential explanation as to how semantic representation might influence the order of lexical-level

utterance plans. When someone hears a word or a sequence of words, activation from that input simultaneously feeds forward to phonological representations and feeds back to semantic representations as well. After initial encoding, lexical activation is determined by repeated interaction with semantic and phonological representation. Serial ordering errors occur when the relative activation levels of the lexical items change because of this interaction. The detrimental effects of phonological similarity on production processes are one well known example of this process (see Dell, 1986). This account also suggests, however, that feed forward input from semantics also influences the activation of elements within a lexical-level utterance plan. If the maintenance of information in verbal WM is achieved by virtue of activation of language-production architecture, this leads to the prediction that disrupting semantic processing should influence the relative activation of lexical-level representations, thus influencing serial ordering.

The present study is designed to test the extent to which semantic representation interacts with lexical and phonological representation to influence serial ordering processes in a test of verbal STM, delayed serial recall. The means by which we test this interaction is through a dual-task paradigm that requires accessing semantic representations. Dual-task distractor procedures have long been employed to interfere with domain specific (Murray & Newman, 1973; Proctor & Fagnani, 1978) and domain general (Kane & Engle, 2000) processing in WM. For instance, in the verbal domain, numerous studies have used the Brown-Peterson paradigm (Brown, 1958; Peterson & Peterson, 1959), or concurrent articulation of an irrelevant syllable (Levy, 1971; Murray, 1968) to disrupt rehearsal and articulation mechanisms. Moreover, evidence of semantic coding in verbal STM was provided by a series of studies by Wickens (see Wickens, 1973, for review), in which changing semantic categories across list material released people from proactive interference in the Brown-Peterson paradigm. To date, however, comparable dual-task procedures have yet to be employed to explore a role for semantic processing in verbal STM (cf. Haarmann & Usher, 2001).

The dual-task procedure in this study requires participants to perform two different picture-judgment tasks as they are engaged in delayed serial recall of auditorily presented material, one requiring a semantic judgment (i.e., is this picture of an animal a dog or not?), the other a visual-perceptual judgment (i.e., is this line oriented up to the left or up to the right?). Orthogonal to the dual-task condition, phonological overlap is also manipulated to determine whether the combination of phonological and semantic interference is particularly detrimental to serial recall performance. Finally, the semantic content of the material being remembered is manipulated through use of concrete or nonword lists. If lexical-level representations are maintained via repeated interaction between semantic, lexical, and phonological representations, as the semantic-binding and language production accounts suggest, then we predict that the picture judgment task will interact with the nature of the list material being used. Specifically, we predict that relative to the line-orientation judgments, simultaneously engaging in semantic categorization will increase the number of serial ordering errors for concrete word lists but will not affect error proportions in nonword lists.

Experiment 1: Delayed Serial Recall of Concrete Word Lists

As a first step toward assessing whether simultaneous performance of a semantic categorization task will affect serial recall performance, we begin with concrete word stimuli that, when heard or read, easily evoke a mental image. If people are using the semantic representations of the concrete words to maintain and/or recall information during verbal STM tasks, then engaging in semantic categorization judgments requiring access to visual semantic representations should create more errors relative to the line orientation judgments that do not.

Method

Participants. Twenty-nine undergraduate students (19 female, 10 male) enrolled in an introductory psychology course at the University of Wisconsin—Madison participated for course credit. Their ages ranged from 18–22 years ($M = 19$, $SD = 0.9$).

Materials.

Words. Forty lists of five concrete words were generated for this experiment. Words were digitally recorded at a sampling rate of 44,100 kHz by an adult male speaking in a monotone intonation. Half of the lists were composed of phonologically overlapping items, which was defined as words sharing a common rhyme unit (e.g., the /æt/ sound of the word *cat*). Lists were matched for relevant psycholinguistic variables (see Table 1), including concreteness/imageability, written frequency (Burgess & Livesay, 1998), number of letters, number of syllables, and average positional and biphone phonotactic probability (Vitevitch & Luce, 2004), which were corrected for stimulus length (Storkel, 2004). The only relevant parameter identified for which overlapping and nonoverlapping lists were not matched on these stimuli was the number of phonemes, where nonoverlapping lists on average had slightly more. This small difference is unlikely to have affected the results reported below, and if it were to have an effect, would be in the direction of reducing the effect of phonological overlap, because the nonoverlapping items are slightly longer.

Concreteness ratings were made by a separate group of participants ($N = 30$; M age $= 20.1$ years, $SD = 1.2$) from those used in this experiment. Participants listened to the words individually and were asked to judge whether the word was abstract or concrete on a scale from 1 (*abstract*) to 7 (*concrete*). Concrete words were defined as those that represent an actual substance or object in the world and therefore more easily evoking a mental image; abstract words were defined as those that do not represent an actual object or substance in the world and therefore do not evoke a mental image. Although this definition conflated concreteness and imageability, it accurately captured the fact that some words have a visual-semantic representation associated with them, whereas others do not.

Pictures. To elicit a range of reaction times and accuracies for making the semantic categorization and line-orientation judgments, a large set of animal pictures were selected and line drawings generated. Separate groups of participants made judgments for each set of stimuli, which were used to select items for the main experiment.

One group of participants ($N = 16$; M age $= 19.1$ years, $SD = 0.9$) made judgments about the orientation of lines of three differ-

Table 1
Descriptive Statistics for Psycholinguistic Variables Across Which Overlapping and Nonoverlapping Lists Were Matched in Each Experiment

List type	Concreteness	Log written frequency	No. of phonemes	No. of syllables	Positional phonotactic probability	Biphone phonotactic probability
			Experiment 1: Concrete words			
Overlap	6.16 (1.69)	3.79 (0.76)	**3.31 (0.63)**	1.04 (0.20)	0.82 (0.27)	0.11 (0.90)
Nonoverlap	6.14 (1.70)	3.79 (0.77)	**3.52 (0.82)**	1.05 (0.22)	0.84 (0.38)	0.04 (0.97)
			Experiment 2: Nonwords			
Overlap			3 (0)	1 (0)	0.68 (0.21)	−0.37 (0.40)
Nonoverlap			3 (0)	1 (0)	0.66 (0.23)	−0.43 (0.41)
			Experiment 3: Within-participants			
Concrete words	6.16 (1.69)	3.79 (0.76)	3.52 (0.82)	1.05 (0.22)	0.84 (0.38)	0.04 (0.97)
Nonwords			3.51 (0.81)	1.09 (0.35)	0.78 (0.17)	0.08 (0.33)

Note. Boldface type indicates a significant difference ($p < .05$).

ent lengths and two different thicknesses. Lines were oriented between $-89°$ to $+89°$ at 1° increments and subtended 4, 9, and 14 degrees of visual angle, respectively. Stimuli were presented until the participant responded. They were followed by a backward mask for 500 ms (a black square subtending 24.5 degrees of visual angle). Participants were instructed to respond with a left key press (the letter *v* on a keyboard) if the stimulus was oriented up toward the left (i.e., all negatively oriented angles) and with a right key press (the letter *b* on a keyboard) if the stimuli were oriented up to the right (i.e., all positively oriented angles). Participants made a total of 534 judgments.

A separate group of participants ($N = 16$; M age $= 19.4$ years, $SD = 1.3$) made judgments about whether a picture of an animal was a dog or not. Pictures were public domain color photographs of a single animal in a natural setting, subtending 17.5 degrees of visual angle. Participants were instructed to respond with a left key press (the letter *v* on a keyboard) if the animal was a dog and with a right key press (the letter *b* on a keyboard) if the animal was not a dog. Pictures were presented one at a time, remained on the screen until participants responded, and were followed by the same backward mask as was used in the line judgments. Participants made a total of 648 judgments.

Following collection of the judgments, 80 line and 80 animal pictures were matched in pairs for speed and accuracy. To norm the stimuli for speed, reaction times for inaccurate responses and for those responses that took longer than 2 s were removed. Mean reaction time for the stimuli was 600 ms ($SD = 38$ ms) for the line and 602 ms ($SD = 25$ ms) for the animal pictures. Mean accuracy was 95.2% ($SD = 4.3\%$) for the line and 96.2% ($SD = 4.1\%$) for the animal stimuli.

Procedure.

Pretesting. Immediate word span was assessed for each participant prior to beginning the experiment. Participants heard a tone indicating the start of a trial followed by a series of words presented at a rate of one per second. At the offset of the last word, another tone was played which served as a cue to recall the list they had just heard in the order in which it was presented. Following two practice trials of lists containing two words, participants tried to recall lists starting with two items up to seven items,

with the list length increasing by one item every two trials. Participants continued this procedure until they failed at both lists of a given length. A participant's span was defined as the last list length at which they had correctly recalled at least one of the lists in the correct serial order. To try to equate the serial recall task for difficulty across individuals, participants were then run at their predetermined span. However, given the need to match overlapping and nonoverlapping lists on the psycholinguistic variables described above, anyone with a span greater than five was run at list lengths of five items. In this experiment, the average word span for individuals was 5.14 items ($SD = 0.74$).

Dual-task. Following pretesting, participants practiced each of the tasks individually prior to practicing them together. Participants began with the picture-judgment tasks. Task instructions and stimulus presentation were the same as in the norming study described above. Participants first made 50 semantic categorization (i.e., dog/not dog) judgments followed by 50 line-orientation judgments. Although this fixed practice order may have added unwanted variance, participants had ample time to practice each task prior to engaging in the dual-task paradigm (see below).

Following practice of the picture-judgment tasks, participants received instructions for the delayed serial recall task. Instructions were the same as in pretesting, except that participants were informed that they would be recalling lists only at their predetermined span and that there would be a brief delay (3 s) between presentation of the words and the recall cue. To maintain the relative order of the list material presented, participants were instructed to say the word "blank" for any item that they could not recall. After practicing the serial recall task, participants practiced performing both the serial recall and picture-judgment tasks at the same time. To ensure that participants were engaged in the picture-judgment task, they were instructed that the picture task was their primary task and that they should try to be as fast and accurate as possible. Both in practice trials and in the actual experiment, participants performed picture judgments alone for 30 s before performing both tasks simultaneously each time a new picture-judgment task started. Given that participants were run at their predetermined span, the duration of the serial recall task varied between 11 s and 17 s. A total of five dual-task practice trials were completed by each participant.

The experiment was conducted in two blocks of 20 trials each, with a break between each block. Each block was divided in half, with participants completing the semantic categorization and line-orientation judgment tasks in each half of the block. The order of the picture tasks was counterbalanced across participants, with half performing in the following order: semantic, line, line, semantic. Participants were instructed about which of the two picture tasks they would be completing before starting each half of the block.

Our reason for using delayed (instead of immediate) serial recall was twofold. First, instituting a delay was designed to maximize the likelihood of observing effects of semantic interference by virtue of repeated interaction of phonological, lexical, and semantic levels of representation. Second, delayed serial recall permits us to vary the timing of the picture-judgment tasks with respect to the different stages of the recall task (i.e., encoding, delay or recall) in future experimentation.

Data scoring.

Serial recall performance. Data scoring for the serial recall task followed methods used in Acheson and MacDonald (2009a). Rather than scoring an item according to whether it was recalled in the correct serial position (i.e., serial recall accuracy) or recalled at all (i.e., item accuracy), participant responses were scored using speech-error analysis. The advantage of this type of scoring is that it provides a more precise classification of the types of errors people made and unambiguously separates item and ordering errors. In this case, an incorrect response was scored according to the linguistic unit over which the error was made (an individual phoneme or a whole item) and for the type of error (transpositions, noncontextual substitutions, omissions, and additions). Following the conventions of Bjork and Healy (1974), items or phonemes were coded as transposition errors if they occurred in the wrong serial position within a list (these errors have also been referred to as contextual substitutions in the production literature; Shattuck-Hufnagel, 1979). Noncontextual substitution errors thus refer to recall of a phoneme or item that was not in the present list. For example, if an individual was trying to recall the list "cat hat bat mat" and recalled the list as "hat *blank* blat rat," we scored the data in the following way. The first response (*hat*) represents a transposition (i.e., an ordering error) and was coded at both the item and the phoneme level, as the response could have been generated either by transposing the item *hat* for *cat* or by substituting the /h/ for the /k/ phonemes. No item was recalled in the second list position (i.e., the participant said *blank*), and this would be scored as an item omission. The third item (*blat*) represents an addition of the phoneme /l/. The last item (*rat*) represents a noncontextual phoneme substitution, as the phoneme /r/ was not present in the list that was presented.

Two points about the scoring bear mention. First, given the ambiguity in whether a transposition of a whole item for another is occurring at the item or phoneme level, all phoneme transposition errors reported below do not include responses that were also coded as an item transposition. Second, it is often the case that when an individual makes a speech error, even when that error is noncontextual, the resulting response is a word (i.e., the lexicality bias; Baars, Motley, & MacKay, 1975; Dell & Reich, 1981). In the present study, a response was scored as a noncontextual, item substitution only if the item had been encountered in a previous list. In this respect, noncontextual, item substitutions corresponded to what WM researchers have termed an *item* or *extralist intrusion*

error (e.g., Henson, 1998). Thus, in the previous example, the erroneously recalled word "rat" would only have been scored as a noncontextual, item substitution if it had previously been encountered by the participant in the experiment.

To conduct the speech-error analysis, all responses were transcribed by two trained individuals using the phonetic alphabet in the *Carnegie Mellon University Pronouncing Dictionary* (http://www.speech.cs.cmu.edu/cgi-bin/cmudict). To establish reliability, six participants were transcribed by both individuals, and their transcriptions agreed 97.3% of the time.

Picture judgments. To assess the extent to which individuals were trading off between the serial recall and picture-judgment tasks, accuracy and reaction times were collected for the picture judgments. All incorrect responses were excluded from the reaction-time estimates. Participant performance was coded according to whether participants were doing the picture judgment alone (i.e., no recall task) or whether they were concurrently engaged in encoding, the delay or recall portion of the delayed serial recall task. This coding allowed us to assess not only whether the two picture-judgment tasks differed but also whether particular portions of the serial recall task were susceptible to disruption from the picture-judgment tasks.

Participant exclusion. One participant was excluded for not being a native English speaker and four were excluded because they failed to complete the task instructions by pressing the wrong keys to respond to the stimuli for significant portions of the experiment. This latter problem occurred because of a combination of an experimental error in coding which buttons could be used to respond coupled with participants pushing the wrong buttons. Thus, the analyses below include 24 participants.

Results

Serial recall performance. Both the phonological overlap and picture-judgment tasks had an effect on the proportion of speech errors, but only those that showed statistically significant effects are considered in the body of this report. The total number of items containing speech errors as well as proportion of times each item contained each type of speech error is included in Table 2. Table 3 includes the inferential statistics resulting from a 2 (Phonological Overlap) × 2 (Picture-Judgment Task) repeated-measures analysis of variance (ANOVA) on the proportion of each type of speech error.

Main effects of phonological overlap were observed for both phoneme and item transpositions. Item transpositions demonstrated a classic phonological similarity effect (Conrad, 1964; Wickelgren, 1965), with more errors for overlapping relative to nonoverlapping lists. Phoneme transpositions showed the reverse effect. Although this latter effect may seem surprising, such reversals have been observed before (see Lian, Karlsen, & Eriksen, 2004) and likely reflect the fact that the vast majority of phoneme transpositions in overlapping lists resulted in production of one of the items in the list. A main effect of the picture-judgment task was observed for a single error type: item transpositions. Individuals produced more item-ordering errors when performing the semantic categorization relative to the line-orientation task. Finally, a Phonological Overlap × Picture Judgment Task interaction was observed for phoneme additions. Post hoc tests (Tukey's Honestly Significant Difference here and throughout the rest of the article)

Table 2

Total Number (Mean Proportion of Items) for Each Type of Speech Error as a Function of Picture-Judgment Task and Phonological Overlap Condition for Concrete Word Lists (Experiment 1)

Speech error type	Dual-task condition	Phonological overlap condition	
		Overlap	Nonoverlap
Transposition			
Phoneme	dog	360 (0.06, 0.08)	420 (0.07, 0.08)
	line	183 (0.03, 0.03)	417 (0.07, 0.06)
Item	dog	1,195 (0.20, 0.09)	680 (0.12, 0.09)
	line	1,090 (0.18, 0.08)	500 (0.09, 0.07)
Noncontextual substitution			
Phoneme	dog	255 (0.04, 0.03)	215 (0.04, 0.03)
	line	240 (0.04, 0.03)	235 (0.04, 0.05)
Item	dog	15 (0.003, 0.007)	20 (0.004, 0.009)
	line	10 (0.004, 0.009)	25 (0.002, 0.006)
Omission			
Phoneme	dog	45 (0.008, 0.01)	10 (0.002, 0.006)
	line	35 (0.006, 0.01)	30 (0.005, 0.01)
Item	dog	805 (0.14, 0.08)	785 (0.15, 0.08)
	line	890 (0.15, 0.08)	790 (0.15, 0.09)
Addition			
Phoneme	dog	5 (0.001, 0.004)	25 (0.004, 0.01)
	line	40 (0.007, 0.01)	10 (0.002, 0.006)

Note. Speech error proportions and standard deviation are noted in parentheses.

showed that while people were significantly less likely to make phoneme additions during semantic categorization judgments relative to line-orientation judgments ($\mu D = 0.006$, $SD = 0.008$) on overlapping lists, no such effect was observed for nonoverlapping lists ($\mu D = -0.003$, $SD = 0.008$).

Picture-judgment performance. To assess any tradeoff between the serial recall and picture-judgment tasks, accuracy and reaction time for each of the picture judgments across different serial recall task phases were examined. To anticipate the results below, there was no evidence of a speed–accuracy tradeoff on the picture-judgment task in this experiment.

Accuracy. Individuals showed only a main effect of the serial recall task phase on the accuracy with which they made the picture judgments. Figure 1 contains the mean accuracy for both the semantic categorization and line-orientation judgment tasks across different phases of the serial recall task (i.e., no recall task, encoding, delay and recall). Error bars in this and all subsequent graphs correspond to 95% confidence intervals based on a pooled estimate of within-participants variance (Masson & Loftus, 2003). Results of a 4 (Serial Recall Task Phase) $\times$ 2 (Picture-Judgment Task) ANOVA revealed a main effect of serial recall task phase, $F(3, 69) = 9.24$, $MSE = 0.001$, $p < .001$, but no main effect of picture-judgment task, $F(1, 23) < 1$, and no interaction between the two variables, $F(3, 69) = 1.1$, $MSE = 0.001$, $p > .35$. Examination of Figure 1 and post hoc tests revealed that participants were less accurate during the delay period relative to not performing the serial recall task ($\mu D = -0.02$, $SD = 0.04$) and encoding ($\mu D = -0.02$, $SD = 0.05$), and the same held for the recall phase relative to not performing the serial recall task ($\mu D = -0.02$, $SD = 0.05$) and relative to encoding ($\mu D = -0.03$, $SD = 0.04$). Given the lack of a significant interaction, however, participants were not selectively trading off accuracy on the picture judgments as a function of doing the serial recall task.

Reaction time. Although participants were not trading off accuracy on the picture-judgment task to perform the serial recall task, there were differences in reaction time between the two picture judgments, which varied as a function of serial recall task phase (see Figure 2). These differences were confirmed in a 4 (Serial Recall Task Phase) $\times$ 2 (Picture-Judgment Task) ANOVA in which there was a significant main effect of serial recall task phase, $F(3, 69) = 39.23$, $MSE = 487472$, $p < .001$, a main effect of picture judgment task, $F(1, 23) = 13.71$, $MSE = 458049$, $p < .001$, and an interaction between the two variables, $F(3, 69) =$

Table 3

Inferential Statistics for 2 (Phonological Overlap) $\times$ 2 (Picture-Judgment Task) Repeated-Measures Analyses of Variance for Different Speech Error Proportions for Concrete Word Lists (Experiment 1)

Speech error type	Phonological overlap	Picture-judgment task	Interaction
Transposition			
Phoneme	**$F(1, 23) = 4.29$, $MSE = 0.014$, $p < .05$**	$F(1, 23) < 1$	$F(1, 23) = 2.30$, $MSE = 0.006$, $p > .10$
Item	**$F(1, 23) = 56.3$, $MSE = 0.18$, $p < .001$**	**$F(1, 23) = 5.90$, $MSE = 0.016$, $p < .03$**	$F(1, 23) < 1$
Noncontextual substitution			
Phoneme	$F(1, 23) < 1$	$F(1, 23) < 1$	$F(1, 23) < 1$
Item	$F(1, 23) < 1$	$F(1, 23) < 1$	$F(1, 23) = 1.76$, $MSE = 0.0001$, $p > .15$
Omission			
Phoneme	$F(1, 23) = 2.26$, $MSE = 0.003$, $p > .10$	$F(1, 23) > 1$	$F(1, 23) = 1.46$, $MSE = 0.002$, $p > .20$
Item	$F(1, 23) < 1$	$F(1, 23) = 1.84$, $MSE = 0.002$, $p > .15$	$F(1, 23) < 1$
Addition			
Phoneme	$F(1, 23) < 1$	$F(1, 23) < 1$	**$F(1, 23) = 5.74$, $MSE = 0.0004$, $p < .05$**

*Note.* Boldface type indicates a significant difference ($p < .05$).
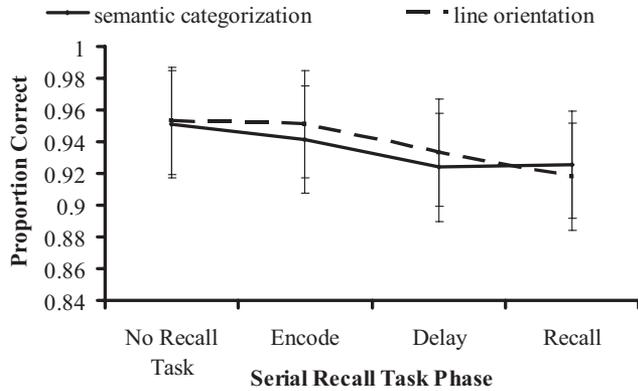
Figure 1. Mean accuracy (95% confidence interval) for the picture-judgment tasks (semantic categorization and line orientation) as a function of serial recall task phase for concrete word lists (Experiment 1).
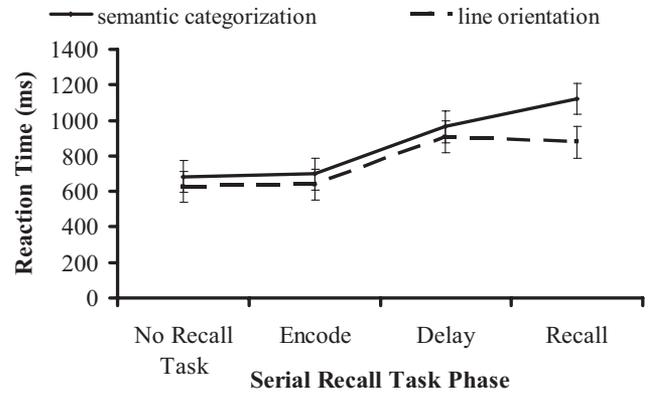
Figure 2. Mean reaction time (95% confidence interval) for the picture-judgment tasks (semantic categorization and line orientation) as a function of serial recall task phase for concrete word lists (Experiment 1).

8.73, $MSE = 21135$, $p < .001$. Examination of Figure 2 and post hoc tests revealed that the main effect of serial recall task phase is explained by the longer reaction time during the delay and recall phases relative to not performing the serial recall task (delay $\mu D = 281$ ms, $SD = 423$; recall $\mu D = 345$ ms, $SD = 323$) and relative to encoding (delay $\mu D = 267$ ms, $SD = 562$; recall $\mu D = 331$ ms, $SD = 363$). The main effect of picture-judgment task is explained by longer reaction time for making the semantic categorization judgment relative to the line-orientation judgments ($\mu D = 105$ ms, $SD = 392$). Finally, post hoc tests on the Serial Recall Task Phase × Picture Judgment Task interaction show a significant difference between the semantic categorization and line-orientation tasks during the recall period ($\mu D = 244$ ms, $SD = 522$) but not during encoding ($\mu D = 60$ ms, $SD = 204$), the delay ($\mu D = 57$ ms, $SD = 722$), or when not engaged in the serial recall task ($\mu D = 59$ ms, $SD = 124$).

Summary and Discussion

The results from Experiment 1 show a clear effect of concurrently performing a semantic categorization task on serial recall performance. Specifically, individuals were more likely to make serial ordering errors (item transpositions) when performing a semantic categorization task relative to a line-orientation task. In addition to this effect on the serial recall task, people also were more likely to slow down on the semantic categorization task relative to the line-orientation task without any difference in accuracy between the two tasks. Such a slowdown is to be expected if the same semantic representations and processes are being used by both tasks, but may also reflect the fact that the semantic categorization decisions are more attention-demanding than the line-orientation judgments. Yet a third possibility is that engaging in the semantic categorization decisions inadvertently activated phonological representations (i.e., people can't help but name the animals in the pictures); hence interference was occurring at a phonological, instead of a semantic, level of representation. Experiment 2 was designed to address these concerns and assess whether the semantic interference effects observed in Experiment 1 are specific to stimuli that contain a semantic representation.

Experiment 2: Delayed Serial Recall of Nonword Lists

Nonwords are legal combinations of speech sounds in a language that do not have a meaning associated with them (e.g., *pof*). Although nonwords share the same phonological properties of real words, they do not benefit from having a lexical-semantic representation; thus an individual's ability to maintain nonwords should be almost entirely reliant on phonological representation. If the semantic categorization decisions are primarily taxing semantic representations and processing, then this task should not affect the serial recall of nonword stimuli.

Method

Participants. Twenty-eight undergraduate students (13 female, 15 male) enrolled in an introductory psychology course at the University of Wisconsin—Madison participated for course credit. Their ages ranged from 18 to 20 years ($M = 19$, $SD = 0.8$).

Stimuli.

Nonwords. The stimuli for the delayed serial recall task were single-syllable, consonant-vowel-consonant nonwords. Half of the lists contained phonological overlap; the other half did not. Overlapping and nonoverlapping lists were matched for phonotactic frequency (see Table 1).

Pictures. The pictures were the same as in Experiment 1.

Procedure. The procedure was the same as in Experiment 1, with the exception that nonwords were used to assess people's immediate memory span in pretesting rather than real words. The average nonword span for participants in this experiment was 3.85 ($SD = 0.60$).

Participant exclusion. A total of four participants were excluded from the analyses, leaving a total of 24. In all four instances, participants failed to press the correct buttons while making the picture judgments during significant portions of the experiment; thus their performance on the picture-judgment task could not be monitored.

Results

Serial recall performance. Results on the serial recall task revealed no effect of the picture-judgment task on the delayed

serial recall of nonwords. Table 4 contains the mean proportion of each of the speech errors in this experiment, and Table 5 the results of the 2 (Phonological Overlap) × 2 (Picture-Judgment Task) repeated-measures ANOVA for each of these speech errors. Results of this ANOVA showed only a main effect of phonological overlap for some of the speech errors, with no main effect of the picture-judgment task, and no interaction. As with the previous experiments, only statistically significant results are discussed here.

Main effects of phonological overlap were observed for three types of speech errors: phoneme and item transpositions and item omissions. A classic phonological similarity effect was observed for phoneme transpositions, with more errors for overlapping relative to nonoverlapping lists, whereas item transpositions and item omissions showed the reverse pattern. These latter results parallel previous research showing a reversal of the phonological similarity effect for item memory in nonword stimuli (Lian et al., 2004).

Picture-judgment performance. Results of the picture-judgment task performance were similar to the previous experiment in which there were differences in accuracy and reaction time between the different STM task phases and differences in reaction time between the different picture-judgment tasks.

Accuracy. As with the previous experiment, only a main effect of serial recall task phase was observed for picture judgment accuracy. Figure 3 contains a graph of the mean accuracy for each picture-judgment task as a function of the different phases of the serial recall task. A 4 (Serial Recall Task Phase) × 2 (Picture-Judgment Task) repeated-measures ANOVA revealed a main effect of serial recall task phase, $F(3, 69) = 13.62$, $MSE = 0.027$,

$p < .001$, but no main effect of picture-judgment task, $F(1, 23) = 1.48$, $MSE = 0.005$, $p > .20$, and no interaction, $F(3, 69) < 1$. As in the previous experiment, the main effect of serial recall task phase is explained by poorer picture-judgment accuracy during the delay and recall portions relative to not performing the serial recall task (both $\mu D = -0.04$, $SD = 0.04$) and relative to the encoding portion (both $\mu D = -0.04$, $SD = 0.05$).

Reaction time. Results for the reaction time to making picture judgments also paralleled the previous experiment with significant main effects of serial recall task phase, $F(3, 69) = 37.56$, $MSE = 537544$, $p < .001$, and picture-judgment task, $F(1, 23) = 45.87$, $MSE = 469929$, $p < .001$, as well as significant interaction between these two variables, $F(3, 69) = 4.19$, $MSE = 5868$, $p < .01$.

Figure 4 contains a graph of the mean reaction time for each picture-judgment task as a function of serial recall task phase. Examination of this figure and post hoc tests reveals that the main effect of serial recall task phase is explained by the longer reaction time during the delay and recall portions of the serial recall trials relative to not doing the serial recall task (delay $\mu D = 222$ ms, $SD = 423$; recall $\mu D = 304$ ms, $SD = 323$) and relative to encoding (delay $\mu D = 196$ ms, $SD = 463$; recall $\mu D = 278$ ms, $SD = 363$). The main effect of picture-judgment task is explained by the overall longer reaction time for making semantic categorization relative to line-orientation judgments ($\mu D = 142$ ms, $SD = 393$). Finally, post hoc tests following the significant Serial Recall Task Phase × Picture-Judgment Task interaction showed that participants were significantly slower to perform semantic categorization relative to line orientation during the delay ($\mu D = 227$ ms, $SD = 352$) and recall ($\mu D = 168$ ms, $SD = 323$) portions of the task, but no such effect was observed during encoding ($\mu D = 97$ ms, $SD = 116$) or when participants were not doing the serial recall task ($\mu D = 74$ ms, $SD = 82$). Thus, similar to the previous experiment, there is no evidence of a speed–accuracy tradeoff in the picture judgments, as participants showed no difference in accuracy and were slower at performing the semantic categorization relative to the line-orientation judgments.

Summary and Discussion

Results of Experiment 2 confirm that the effect of the semantic categorization task on serial ordering of stimuli in verbal STM is specific to material that contains a semantic representation (i.e., concrete words). In the present case, no differences on the mean proportions of speech errors in delayed serial recall were observed for nonword stimuli between the two different picture judgments. This result is particularly important to the present study, as it demonstrates that our semantic categorization task was primarily disrupting semantic and not phonological processing during performance of the serial recall task. In Experiment 3, we directly compare these effects by manipulating stimulus type within participants.

Experiment 3: Delayed Serial Recall of Words and Nonwords using a Within-Participants Design

The results of the first two experiments suggest that concurrent performance of a semantic categorization task has an effect on the delayed serial recall of material containing semantic content (i.e.,

Table 4
Total Number (Mean Proportion of Items) for Each Type of Speech Error as a Function of Picture-Judgment Task and Phonological Overlap Condition for Nonword Lists (Experiment 2)

Speech error type	Dual-task condition	Phonological overlap condition	
		Overlap	Nonoverlap
Transposition			
Phoneme	dog	985 (0.23, 0.12)	575 (0.16, 0.12)
	line	865 (0.20, 0.15)	640 (0.16, 0.10)
Item	dog	145 (0.03, 0.04)	425 (0.09, 0.06)
	line	185 (0.04, 0.04)	425 (0.09, 0.06)
Noncontextual substitution			
Phoneme	dog	550 (0.12, 0.07)	455 (0.10, 0.08)
	line	465 (0.10, 0.11)	495 (0.11, 0.08)
Item	dog	0	0
	line	0	0
Omission			
Phoneme	dog	20 (0.005, 0.02)	5 (0.001, 0.005)
	line	5 (0.001, 0.005)	5 (0.001, 0.005)
Item	dog	200 (0.04, 0.06)	330 (0.07, 0.07)
	line	180 (0.04, 0.08)	275 (0.06, 0.08)
Addition			
Phoneme	dog	20 (0.005, 0.02)	30 (0.006, 0.02)
	line	35 (0.008, 0.03)	40 (0.009, 0.03)

Note. Speech error proportions and standard deviation are noted in parentheses.

Table 5
Inferential Statistics for 2 (Phonological Overlap) × 2 (Picture-Judgment Task) Repeated-Measures Analyses of Variance for Different Speech Error Proportions for Concrete Word Lists (Experiment 2)

Speech error type	Phonological overlap	Picture-judgment task	Interaction
Transposition			
Phoneme	$F(1, 23) = 75.31, MSE = 0.34,$ $p < .0001$	$F(1, 23) < 1$	$F(1, 23) < 1$
Item	$F(1, 23) = 76.31, MSE = 1.38,$ $p < .0001$	$F(1, 23) = 2.31, MSE = 0.007,$ $p > .1$	$F(1, 23) < 1$
Noncontextual substitution			
Phoneme	$F(1, 23) < 1$	$F(1, 23) < 1$	$F(1, 23) = 1.52, p > .20$
Item	0	0	0
Omission			
Phoneme	$F(1, 23) < 1$	$F(1, 23) < 1$	$F(1, 23) < 1$
Item	$F(1, 23) = 13.69, MSE = 0.02,$ $p < .001$	$F(1, 23) = 1.03, MSE = 0.0014,$ $p > .3$	$F(1, 23) = 1.45, MSE = 0.002,$ $p > .20$
Addition			
Phoneme	$F(1, 23) < 1$	$F(1, 23) < 1$	$F(1, 23) < 1$

Note. Boldface type indicates a significant difference ($p < .05$).

concrete words) but not material that does not (nonwords). Specifically, the semantic categorization decisions led to more serial ordering errors relative to line-orientation judgments for concrete words but not nonwords. Directly comparing this effect across these two experiments is difficult, however, given that participants were not randomly assigned to word lists, and the stimuli were not perfectly matched across all relevant phonological factors. Furthermore, people were run at different list lengths within each experiment, which may have contributed additional, unwanted variance to performance.

To rectify these differences and directly compare the effects of the semantic categorization on serial recall performance, we conducted a third experiment in which semantic content of the list material was varied within participants using a constant list length for each type of material. Unlike the previous two experiments, word and nonword stimuli were matched for all phonological factors (see Table 1). Finally, except for a small effect on phoneme additions in Experiment 1, the picture-judgment task did not interact with phonological overlap manipulation in the first two experiments. Because of this and a desire to keep the experiment to a reasonable length, the overlapping list condition was removed, and participants recalled only nonoverlapping lists.

Method

Participants. Twenty undergraduate students (13 female, 7 male) enrolled in an introductory psychology course at the University of Wisconsin—Madison participated for course credit. Their ages ranged from 18 to 20 years ($M = 18.9$, $SD = 0.8$).

Stimuli. Stimuli consisted of 40 lists of single and multisyllabic concrete words and nonwords (20 each). Stimuli within a list did not contain phonological overlap and were matched for relevant psycholinguistic factors (see Table 1).

Pictures. The pictures used were the same as Experiments 1 and 2.

Procedure. The procedure was similar to Experiments 1 and 2, except that no pretesting was used. On the basis of the average span of participants in the first two experiments, participants were
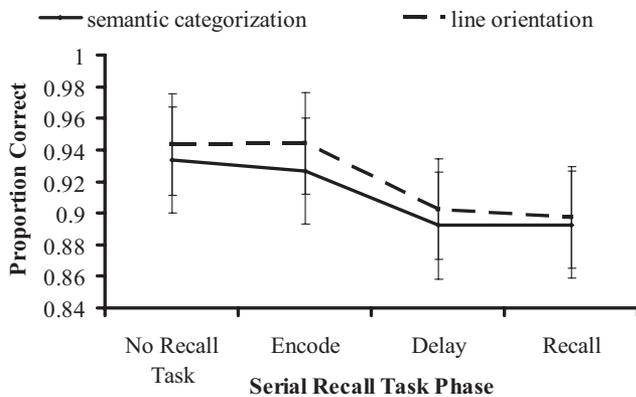
Figure 3. Mean accuracy (95% confidence interval) for the picture-judgment tasks (semantic categorization and line orientation) as a function of serial recall task phase for nonword lists (Experiment 2).
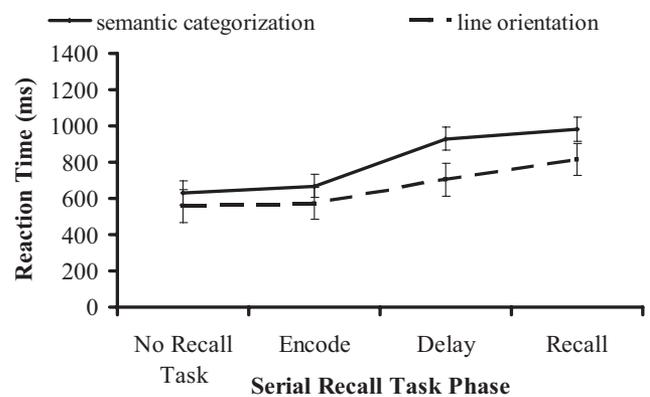
Figure 4. Mean reaction time (95% confidence interval) for the picture-judgment tasks (semantic categorization and line orientation) as a function of serial recall task phase for nonword lists (Experiment 2).

run at list lengths of five items for concrete words and four items for the nonwords. The difference in list length between words and nonwords was done to try to equate task difficulty across stimulus conditions. Stimuli were presented in blocks of 10 lists, and the order of blocks and picture judgments was counterbalanced across participants.

Results

Serial recall performance. Results confirmed the differential effects of the picture-judgment tasks on word and nonword stimuli observed in Experiments 1 and 2. Table 6 contains the total number and mean proportion of the speech errors in this experiment, and Table 7 the results of a 2 (Stimulus Type) × 2 (Picture-Judgment Task) repeated-measures ANOVA for each of these error types. Main effects of stimulus type were observed for phoneme transpositions, phoneme and item omissions, as well as phoneme additions. Main effects of picture-judgment task and a Stimulus Type × Picture-Judgment Task interaction were observed for item transpositions.

Examination of Table 6 shows that the main effect of stimulus type for phoneme transpositions, omissions, and additions is driven by the higher numbers of these errors for nonwords relative to concrete words. More item omissions, however, were observed for concrete words relative to nonwords. The main effect of picture-judgment task reflects the fact that more item transpositions were observed overall for the semantic categorization relative to the line-orientation judgments; however, this main effect is driven by the results of the interaction. Post hoc tests showed that there were significantly more item transpositions for semantic categorization judgments relative to line-orientation judgments for

concrete words ($\mu_D = 0.05$, $SD = 0.31$), but no such effect was observed for nonwords ($\mu_D = -0.01$, $SD = 0.25$). Thus, similar to the results of the previous two experiments, the effect of semantic categorization decisions on item-ordering errors in Experiment 3 was specific to material containing a semantic representation.

Picture-judgment performance. Results of the picture-judgment performance mirrored those of the first two experiments, with no evidence of any differences between the two different stimulus types. As before, there was no evidence of a speed–accuracy tradeoff, although participants were slower to make semantic categorization decisions relative to line-orientation judgments.

Accuracy. Figure 5 contains a graph of mean accuracy as a function of serial recall task phase, stimulus type, and picture-judgment task, and Table 8 shows the results of a 2 (Picture-Judgment Task) × 2 (Stimulus Type) × 4 (Serial Recall Task Phase) repeated-measures ANOVA. Similar to the previous two experiments, the results of the ANOVA showed only a main effect of serial recall task phase. Examination of Figure 5 and post hoc tests showed that the main effect of serial recall task phase was driven by the fact that individuals were significantly less accurate at making picture judgments during the delay and recall phases relative to not being engaged in the serial recall task (delay $\mu_D = -0.03$, $SD = 0.08$; recall $\mu_D = -0.03$, $SD = 0.08$) and relative to encoding (delay $\mu_D = -0.03$, $SD = 0.08$; recall $\mu_D = -0.03$, $SD = 0.09$). Given the lack of significant interactions with stimulus types, these results did not vary as a function of whether participants were remembering words or nonwords.

Reaction time. Figure 6 contains a graph of mean reaction time as a function of serial recall task phase, stimulus type, and picture-judgment task, and Table 8 contains the results of a 2 (Picture-Judgment Task) × 2 (Stimulus Type) × 4 (Serial Recall Task Phase) repeated-measures ANOVA. Results of the ANOVA revealed main effects of the picture-judgment task and serial recall task phase, as well as an interaction between these two variables. Overall, participants were slower at making semantic categorization decisions relative to line-orientation judgments ($\mu_D = 129.2$ ms, $SD = 160$). The main effect of serial recall task phase came from the fact that participants were significantly slower at making picture judgments during the delay and recall phases relative to not doing the serial recall task (delay $\mu_D = 96$ ms, $SD = 143$; recall $\mu_D = 172$ ms, $SD = 150$) and relative to encoding (delay $\mu_D = 94$ ms, $SD = 156$; recall $\mu_D = 169$ ms, $SD = 164$). Finally, the interaction of Picture-Judgment Task × Serial Recall Task Phase came from the significant differences in reaction time between the semantic categorization and line-orientation judgments during the delay ($\mu_D = 172$ ms, $SD = 168$) and recall ($\mu_D = 147$ ms, $SD = 195$) phases but not during encoding ($\mu_D = 104$ ms, $SD = 107$) or when participants were not engaged in the serial recall task ($\mu_D = 91$ ms, $SD = 81$). Contrary to Experiment 1, some differences in reaction time were observed for concrete words during the delay period. Despite this, the results generally parallel those observed from the previous two experiments and, it is important to note, show no evidence of differences between the different stimuli.

General Discussion

Three experiments explored the influence of lexical-semantic and phonological factors on the maintenance of information in verbal

Table 6

Total Number (Mean Proportion of Items) for Each Type of Speech Error as a Function of Picture-Judgment Task and Stimulus Type (Experiment 3)

Speech error type	Dual-task condition	Stimulus type	
		Nonwords	Concrete words
Transposition			
Phoneme	dog	282 (0.30, 0.46)	158 (0.18, 0.38)
	line	305 (0.32, 0.47)	174 (0.19, 0.39)
Item	dog	67 (0.07, 0.25)	133 (0.14, 0.33)
	line	75 (0.08, 0.26)	93 (0.09, 0.28)
Noncontextual substitution			
Phoneme	dog	117 (0.12, 0.32)	38 (0.04, 0.19)
	line	120 (0.12, 0.32)	51 (0.05, 0.22)
Item	dog	9 (0.01, 0.09)	14 (0.01, 0.12)
	line	12 (0.02, 0.11)	4 (0.004, 0.06)
Omission			
Phoneme	dog	17 (0.02, 0.13)	9 (0.01, 0.09)
	line	18 (0.02, 0.13)	8 (0.01, 0.09)
Item	dog	45 (0.05, 0.21)	146 (0.15, 0.35)
	line	46 (0.05, 0.21)	123 (0.12, 0.33)
Addition			
Phoneme	dog	13 (0.01, 0.11)	3 (0.003, 0.05)
	line	18 (0.02, 0.13)	4 (0.004, 0.06)

Note. Speech error proportions and standard deviation are noted in parentheses.

Table 7

Inferential Statistics for 2 (Stimulus Type) × 2 (Picture-Judgment Task) Repeated-Measures Analyses of Variance for Different Speech Error Proportions in Experiment 3

Speech error type	Stimulus	Picture-judgment task	Interaction
Transposition			
Phoneme	**$F(1, 19) = 62.85, MSE = 0.36, p < .001$**	$F(1, 19) < 1$	$F(1, 19) < 1$
Item	$F(1, 19) = 2.56, MSE = 0.005, p > .1$	**$F(1, 19) = 7.75, MSE = 0.046, p < .02$**	**$F(1, 19) = 11.83, MSE = 0.013, p < .01$**
Noncontextual substitution			
Phoneme	$F(1, 19) = 76.23, MSE = 0.17, p < .001$	$F(1, 19) < 1$	$F(1, 19) < 1$
Item	$F(1, 19) < 1$	$F(1, 19) = 1.35, MSE = 0.0004, p > .25$	$F(1, 19) = 3.96, MSE = 0.0013, p > .05$
Omission			
Phoneme	**$F(1, 19) = 5.26, MSE = 0.0013, p < .05$**	$F(1, 19) < 1$	$F(1, 19) = 1.79, MSE = 0.0006, p > .15$
Item	**$F(1, 19) = 31.11, MSE = 0.25, p < .001$**	$F(1, 19) = 1.98, MSE = 0.006, p > .15$	$F(1, 19) = 1.37, MSE = 0.002, p > .25$
Addition			
Phoneme	**$F(1, 19) = 8.26, MSE = 0.003, p < .01$**	$F(1, 19) = 2.09, MSE = 0.0005, p > .15$	$F(1, 10) < 1$

Note. Boldface type indicates a significant difference ($p < .05$).

STM through use of a novel dual-task paradigm. In the first two experiments, phonological overlap and concurrent performance of two different picture-judgment tasks was manipulated within each experiment, whereas the type of material being remembered (words vs. nonwords) was varied across experiments. In a third experiment, the effect of the picture-judgment tasks on memory for words and nonwords was compared directly in a within-subjects design. Results of Experiments 1 and 2 showed an effect of phonological overlap that mirrored those of previous studies, with more item-ordering errors for word lists (Coltheart, 1993; Henson, Norris, Page, & Baddeley, 1996) and more phoneme-ordering errors for nonword lists (e.g., Acheson & MacDonald, 2009a; Treiman & Danis, 1988). On top of the phonological similarity effect observed across list material, there was a main effect of the picture-judgment task manipulation for concrete words lists (Experiments 1 and 3), which was due to the higher proportion of item-ordering errors for

the semantic categorization relative to line-orientation judgments. Critically, no such effect was found for nonword lists (Experiments 2 and 3), and the magnitude of this effect did not vary as a function of phonological overlap. In sum, the effect of performing a concurrent visual semantic categorization task on serial ordering of information in serial recall performance was specific to list material that contained a visual semantic representation (i.e., concrete words). Although the magnitude of the phonological similarity effects likely varied between the different list material (see Gupta, Lipinski, & Aktunc, 2005, for an excellent review), the following discussion focuses on the effects of concurrently engaging semantic processing on the maintenance of information in verbal STM.

There are two relatively novel aspects of the present investigation. The first is the method by which we examined the influence of semantic processing on serial recall performance. Although
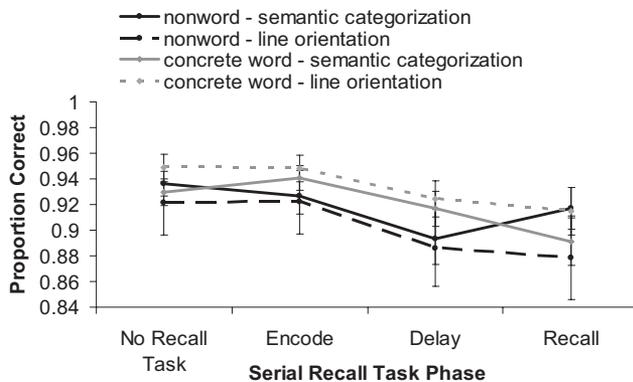
Figure 5. Mean accuracy (95% confidence interval) for the picture-judgment tasks (semantic categorization and line orientation) as a function of stimulus type (concrete words and nonwords) and serial recall task phase (Experiment 3).
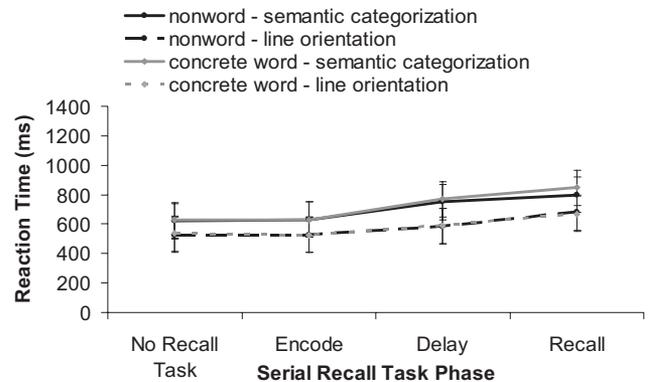
Figure 6. Mean reaction time (95% confidence interval) for the picture-judgment tasks (semantic categorization and line orientation) as a function of stimulus type (concrete words and nonwords) and serial recall task phase (Experiment 3).

Table 8

Inferential Statistics for 2 (Picture-Judgment Task) × 2 (Stimulus Type) × 4 (Serial Recall Task Phase) Repeated-Measures Analyses of Variance for Accuracy and Reaction Time in the Picture-Judgment Tasks in Experiment 3

Conditions and interactions	Accuracy	Reaction time
Picture-judgment task (P)	$F(1, 19) < 1$	$\boldsymbol{F(1, 19) = 42.78, MSE = 1237267, p < .001}$
Stimulus (S)	$F(1, 19) = 1.96, MSE = 0.02, p > .15$	$F(1, 19) < 1$
Working memory task phase (WM)	$\boldsymbol{F(3, 57) = 10.53, MSE = 0.027, p < .001}$	$\boldsymbol{F(3, 57) = 29.85, MSE = 544433, p < .001}$
P × S	$F(1, 19) = 1.48, MSE = 0.02, p > .20$	$F(1, 19) < 1$
P × WM	$F(3, 57) < 1$	$\boldsymbol{F(3, 57) = 9.80, MSE = 29164, p < .001}$
S × WM	$F(3, 57) = 1.57, MSE = 0.003, p > .20$	$F(3, 57) < 1$
P × S × WM	$F(3, 57) = 1.83, MSE = 0.003, p > .15$	$F(3, 57) = 1.02, MSE = 3569, p > .35$

Note. Boldface type indicates a significant difference ($p < .05$).

dual-task paradigms have long been used to disrupt rehearsal (Brown, 1958; Peterson & Peterson, 1959) or to study how articulatory and phonological representation interact in STM (Levy, 1971; Murray, 1968), few attempts have been made to engage semantic processing simultaneously during verbal STM task performance (e.g., Haarmann & Usher, 2001). Thus, the novelty of the present results lies not simply in the fact that performing a secondary task can disrupt STM performance but, rather, in the specificity with which the semantic categorization task had its effect. There are important theoretical implications raised by the fact that engaging semantic processing specifically affected memory for the serial order of concrete word lists and not simply memory for the items themselves (e.g., Poirier & Saint-Aubin, 1995). To our knowledge, there are only a handful of studies that have shown that lexical-semantic manipulations can specifically influence serial ordering processes (Acheson et al., 2010; Baddeley, 1966; Jefferies, Frankish, & Lambon Ralph, 2006a; Murdock & Vom Saal, 1967; Romani et al., 2008). Such results do not fit neatly within classic conceptualizations of verbal WM maintenance that explicitly dissociate short-term phonological and long-term semantic codes and their respective roles in serial order and item memory (e.g., Baddeley, 1986, 2007). The present results can, however, be accommodated within linguistically motivated frameworks of verbal WM maintenance.

One such framework is provided by researchers who have suggested that the maintenance of information in verbal WM might be accomplished via temporary activation of representations with the language production architecture (Acheson & MacDonald, 2009b; Saffran & Martin, 1997). Page et al. (2007), for instance, likened maintenance in verbal WM to the maintenance of a lexical-level plan of the sort that is used in normal language production. Critically, within models of language production, the representation of a lexical item contains conceptual information (i.e., lemmas; see Levelt et al., 1999) and, in some connectionist models, phonological information as well (e.g., Plaut & Kello, 1999). Although typical WM tasks differ from normal language production to the extent that the "message" is dictated by the researcher, once a lexical-level plan has been generated, the same processes may be involved. Interactive activation accounts of language production, for instance, would predict that relative activation of lexical items will be influenced by both semantic and phonological representations (Dell & O'Seaghdha, 1992; Foygel & Dell, 2000). According to this type of account, the semantic categorization task in the present experiment may have interfered with the feed-

forward connections from semantic to lexical-level representations, potentially altering their relative activation levels, thus rendering them more susceptible to movement within a list.

Such an account is similar to the semantic-binding hypothesis, which suggests that semantic representations can influence the binding of sublexical phonological representation (Patterson et al., 1994), but differs to the extent that it emphasizes the binding of whole lexical units to relative positions within a list. This being said, an interactive activation account would also predict that disruption of lexical-semantic representation should make people more reliant on the phonological characteristics of the list material (e.g., Dell, Schwartz, Martin, Saffran, & Gagnon, 1997). In relation to the experiments presented here, the effect of the picture-judgment task on ordering errors would be expected to be larger when lists were phonologically similar relative to when they were dissimilar. Some support of this was provided in our data by phoneme-ordering errors in concrete word lists, in that the difference between the two picture-judgment tasks (Semantic Categorization − Line Orientation) was larger for phonologically similar (2.5%) relative to dissimilar lists (−0.8%). However, given that no statistically significant interaction on serial ordering errors was observed between the picture-judgment task and the phonological overlap manipulations, this evidence should be interpreted with caution.

Given the emphasis we have placed on linguistically motivated accounts above, an important question to ask is why there was little if any evidence of an interaction between the semantic and phonological manipulations on sublexical or whole item serial ordering in the present study. We think the answer to this is threefold. The first is that the semantic manipulation in the present study was likely too weak to "unglue" the semantic binding on phonological representations in typical young adults, as has been the case in patient populations with semantic impairments (Knott et al., 2000; Patterson et al., 1994). The overwhelming majority of ordering errors made by unimpaired adults occurs over entire items (Henson, 1998), and only under testing conditions where semantics are extremely impoverished (e.g., when there are high numbers of nonwords; Acheson & MacDonald, 2009a; Jefferies et al., 2006a; Treiman & Danis, 1988) are a preponderance of such errors observed. The present results may also speak to the fact that a distinct lexical level of representation exists between semantic and phonological representation (as instantiated in many models of language production; Dell, 1986; Levelt et al., 1999) and, more speculatively, that the feed-forward connections from semantics

influence lexical representations more than feedback from phonological levels of representation. Finally, although it is quite possible that phoneme transpositions were occurring, a number of these transpositions would have resulted in production of a nonword. Given that participants were aware of the material they were remembering, such nonword responses may have been strategically corrected by participants prior to responding in the context of word lists (see Jefferies, Frankish, & Noble, 2009). In fact, the combination of this (more phoneme-transposition errors induced by semantic categorization) coupled with strategic editing of spoken responses may be expected to lead to more whole item-transposition errors, as was observed in the present study.

An alternative to the more language-motivated accounts discussed above is that the semantic categorization task may have interfered with the process of trace redintegration during recall (e.g., Hulme et al., 1997; Schweickert, 1993). According to this account, phonological representations that have decayed or have been interfered with during maintenance are refreshed through a comparison to long-term, linguistic representation at the time of recall. This account conflicts with the present results, however, in that the redintegration process is generally thought to occur on an item-by-item basis (hence, it should not affect serial ordering). However, Stuart and Hulme (2000) have suggested that long-term memory may include a representation of interitem associations, which could in turn influence serial ordering processes through a redintegration-like mechanism. It is possible, then, that the effect of the semantic categorization task was to reduce people's ability to access these interitem associations during recall, and this accounts for the serial ordering errors observed in the present study.[1]

Although we think it unlikely that the material in our lists had strong interitem associations in long-term memory, a consideration of trace redintegration accounts highlights what is, at present, one of the central questions to verbal WM research: At what stage of processing (e.g., maintenance, recall) is lexical-semantic information affecting performance? Resolution of this question speaks to an even broader issue of whether short- and long-term memory are dissociable entities, a debate that has been ongoing for many years (e.g., Atkinson & Shiffrin, 1968; Crowder, 1993). As is discussed in the following, one of the advantages of the present methodology is that it can be used to test directly whether engaging semantic processing primarily affects maintenance or recall processes. Before getting to this discussion, however, there are some limitations to the present results that must be considered.

Limitations

The first limitation of our new paradigm is that the reaction-time data on the picture-judgment task suggest that, on some levels, the line-orientation and semantic categorization tasks may not have been perfectly matched in difficulty in the context of the dual-task paradigm. Although stimuli were matched for speed and accuracy before being used concurrently with delayed serial recall, individuals were systematically slower at making the semantic categorization relative to the line-orientation judgments. Although such slowing might be expected if both the serial recall and semantic categorization tasks were drawing on the same resources, the fact that there was an equivalent slowdown for word and nonword stimuli might be taken as evidence that the semantic task was simply more difficult or attention demanding than the line-orientation task. It is noteworthy, for instance, that participants were slower to make the semantic categorization judgments relative to line-orientation judgments specifically during the most attention-demanding portions of the experiment: maintenance and recall. However, although this account can explain why there was more of a slowdown for the semantic categorization relative to line-orientation judgments, it cannot explain the specificity of the effect of concurrent semantic categorization on the serial recall of concrete words.

Another possibility that might explain differences in reaction time between the semantic categorization and line-orientation tasks is that the former may have interfered with phonological processing by virtue of the automaticity with which the animal pictures were named (i.e., people cannot avoid saying "cat" when they see a picture of a cat). Although we cannot exclude this possibility, we view it as unlikely for two reasons. First, outside of a small effect on phoneme additions in Experiment 1, there was no interaction between the picture-judgment task and the phonological overlap manipulations. Second, the effect of the semantic categorization task was specific to material that contained a semantic representation. If phonological representations had also been activated, one should have expected effects on nonword stimuli as well. Nonetheless, future research might be directed at finding a line-orientation or another nonsemantic, perceptual task that is better matched for speed and accuracy than the material used in the present experiment.

The second limitation to the present results is that the effect of the dual task may simply be capitalizing on strategies participants were using to maintain the material (e.g., visual imagery). Campoy and Baddeley (2008), for instance, demonstrated that the magnitude of the phonological similarity and word-length effects were dramatically reduced when participants were instructed to use a semantic strategy to encode the material (see also Hanley & Bakopoulou, 2003). From this, one might conjecture that the semantic categorization task in the present study may have had its effect simply because some individuals chose to visualize the list material. Although it is beyond the scope of the present work to address the role of strategy in verbal STM performance, it should be noted that in studies that have investigated this question (e.g., Logie, Della Sala, Laiacona, Chalmers, & Wynn, 1996), less than

---

[1] A third possibility is that the semantic categorization task was influencing control over semantic representations instead of the representations themselves. Jefferies, Hoffman, Jones, and Lambon Ralph (2008) directly compared the error patterns of patients with semantic dementia and transcortical sensory aphasia, the latter of which is associated with deficits in control over semantic representations (Jefferies & Lambon Ralph, 2006). Results of this comparison revealed that whereas semantic dementia patients made primarily sublexical serial ordering errors, those with transcortical sensory aphasia were more likely to produce whole item serial ordering errors. Similarly, Hamilton and Martin (2007) hypothesized that impaired performance in stroke patient ML, who does not show a recall advantage for words over nonwords and is susceptible to numerous item intrusions, is due to a deficit in inhibitory control. Providing one can interpret error patterns as reflecting underlying processes (a potentially problematic endeavor; see Dell & Reich, 1981), the effects of the semantic categorization task in the present study may be attributable to a disruption of some form of control over activated semantic representations in memory.

10% of respondents reported using a visual semantic strategy. In the present investigation, over 70% of our participants in Experiment 1, and 85% of the participants in Experiment 3 showed a higher proportion of ordering errors in the semantic categorization relative to the line-orientation task. To the extent that the strategies reported in Logie et al. (1996) are indicative of strategy use more generally, the present results suggest that the effects of the semantic categorization task on performance were not simply due to strategy.

Finally, it should be noted that different list lengths were used for concrete and nonword lists when the two stimuli were directly compared in Experiment 3. Thus, it is possible that some of the differences we observe between different list types may be due to differences in list length. Although we believe it to be unlikely that this confound was responsible for the pattern of results observed in Experiment 3, future experimentation should be used to address whether the effects of the secondary tasks vary as a function of different list lengths.

Implications and Future Research Directions

One of the benefits of the dual-task methodology used in the present study is that it could potentially be used to test hypotheses about what role semantic representation plays in the maintenance of information in verbal WM. A fundamental debate in memory research is the extent to which WM mechanisms are independent of long-term memory (Baddeley, 2007; Crowder, 1993). One of the most recent instantiations of this debate are questions about when long-term, linguistic knowledge (e.g., concreteness, frequency) influences task performance. Many linguistically motivated accounts posit that these levels of linguistic representations are actively maintained in verbal WM (Acheson & MacDonald, 2009b; Haarmann & Usher, 2001; Martin, et al., 1999; Saffran & Martin, 1997). Trace redintegration accounts, on the other hand, preserve a distinction in which maintenance is purely phonological (e.g., Baddeley, 1986, 2007), and long-term, linguistic knowledge affects performance at the time of recall (Hulme et al., 1997; Schweickert, 1993). The design of the present study does not delineate during which stage of serial-recall performance the semantic categorization task was having its effect. However, the present method could be used to adjudicate between these two accounts by varying the time of serial recall task performance during which people are simultaneously engaging semantic processing (e.g., during encoding, maintenance, or recall only). Whereas linguistically motivated accounts would predict that engaging semantic processing during the delay would affect task performance, trace redintegration accounts would not.

The broader implication of the present investigation is that it may no longer be tenable to assume that the maintenance of information in verbal WM occurs over strictly phonological levels of representation. Although it is likely that a clean-up process such as trace redintegration is occurring, given the automaticity with which words activate their respective semantics, it is also likely that such representations are actively maintained in WM. Although we have focused on phonological, lexical, and semantic levels of representation in the present study, the full picture of verbal WM is likely to include maintenance across all levels of linguistic representation, as there are multiple means by which any word can be coded in memory (Wickens, 1973). In the case of verbal WM,

the vast majority of research has been conducted using stimuli and testing procedures (e.g., repeatedly sampling from a closed set of letters, digits, etc.) that strongly encourage the use of phonological, acoustic and articulatory codes. This has led to a common assumption that such coding is therefore the sole level of representation over which information is maintained in verbal WM. The present results not only speak against this assumption but speak to what we believe to be a larger truth about WM as well: People will use as many different representational codes in WM as are permitted by the stimuli used and the testing conditions employed.

References

Acheson, D. J., & MacDonald, M. C. (2009a). Twisting tongues and memories: Explorations into the relationship between language production and verbal working memory. *Journal of Memory and Language, 60,* 329–350. doi:10.1016/j.jml.2008.12.002

Acheson, D. J., & MacDonald, M. C. (2009b). Verbal working memory and language production: Common approaches to the serial ordering of verbal information. *Psychological Bulletin, 135,* 50–68. doi:10.1037/a0014411

Acheson, D. J., Postle, B. R., & MacDonald, M. C. (2010). The interaction of concreteness and phonological similarity in verbal working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36,* 17–36. doi:10.1037/a0017679

Allen, R., & Hulme, C. (2006). Speech and language processing mechanisms in verbal serial recall. *Journal of Memory and Language, 55,* 64–88. doi:10.1016/j.jml.2006.02.002

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation: II* (pp. 89–195). New York, NY: Academic Press. doi:10.1016/S0079-7421(08)60422-3

Baars, B. J., Motley, M. T., & MacKay, D. G. (1975). Output editing for lexical status in artificially elicited slips of the tongue. *Journal of Verbal Learning and Verbal Behavior, 14,* 382–391. doi:10.1016/S0022-5371(75)80017-X

Baddeley, A. D. (1966). Short-term memory for word sequences as a function of acoustic, semantic and formal similarity. *The Quarterly Journal of Experimental Psychology, 18,* 362–365. doi:10.1080/14640746608400055

Baddeley, A. D. (1972). Retrieval rules and semantic coding in short-term memory. *Psychological Bulletin, 78,* 379–385. doi:10.1037/h0033477

Baddeley, A. D. (1986). *Working memory.* New York, NY: Clarendon Press/Oxford University Press.

Baddeley, A. D. (2007). *Working memory, thought, and action.* New York, NY: Oxford University Press.

Bjork, E. L., & Healy, A. F. (1974). Short-term order and item retention. *Journal of Verbal Learning and Verbal Behavior, 13,* 80–97. doi:10.1016/S0022-5371(74)80033-2

Bourassa, D. C., & Besner, D. (1994). Beyond the articulatory loop: A semantic contribution to serial order recall of subspan lists. *Psychonomic Bulletin & Review, 1,* 122–125.

Brener, R. (1940). An experimental investigation of memory span. *Journal of Experimental Psychology, 26,* 467–482. doi:10.1037/h0061096

Brown, J. (1958). Some tests of the decay theory of immediate memory. *Quarterly Journal of Experimental Psychology, 10,* 12–21. doi:10.1080/17470215808416249

Burgess, C., & Livesay, K. (1998). The effect of corpus size in predicting reaction time in a basic word recognition task: Moving on from Kucera and Francis. *Behavior Research Methods, Instruments & Computers, 30,* 272–277.

Campoy, G., & Baddeley, A. D. (2008). Phonological and semantic strategies in immediate serial recall. *Memory, 16,* 329–340. doi:10.1080/09658210701867302

Caza, N., Belleville, S., & Gilbert, B. (2002). How loss of meaning with preservation of phonological word form affects immediate serial recall performance: A linguistic account. *Neurocase, 8,* 255–273. doi:10.1076/neur.8.3.255.16196

Coltheart, V. (1993). Effects of phonological similarity and concurrent irrelevant articulation on short-term-memory recall of repeated and novel word lists. *Memory & Cognition, 21,* 539–545.

Conrad, R. (1964). Acoustic confusions in immediate memory. *British Journal of Psychology, 55,* 75–84.

Conrad, R., & Hull, A. J. (1964). Information, acoustic confusion and memory span. *British Journal of Psychology, 55,* 429–432.

Cowan, N. (1995). *Attention and memory: An integrated framework.* New York, NY: Oxford University Press.

Crowder, R. G. (1989). Modularity and dissociations in memory systems. In H. L. Roediger, III & F. I. M. Craik (Eds.), *Varieties of memory and consciousness: Essays in honour of Endel Tulving* (pp. 271–294). Hillsdale, NJ: Erlbaum.

Crowder, R. G. (1993). Short-term memory: Where do we stand? *Memory & Cognition, 21,* 142–145.

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review, 93,* 283–321. doi:10.1037/0033-295X.93.3.283

Dell, G. S., & O'Seaghdha, P. G. (1992). Stages of lexical access in language production. *Cognition, 42,* 287–314. doi:10.1016/0010-0277(92)90046-K

Dell, G. S., & Reich, P. A. (1981). Stages in sentence production: An analysis of speech error data. *Journal of Verbal Learning and Verbal Behavior, 20,* 611–629. doi:10.1016/S0022-5371(81)90202-4

Dell, G. S., Schwartz, M. F., Martin, N., Saffran, E. M., & Gagnon, D. A. (1997). Lexical access in aphasic and nonaphasic speakers. *Psychological Review, 104,* 801–838. doi:10.1037/0033-295X.104.4.801

Fallon, B., Groves, K., & Tehan, G. (1999). Phonological similarity and trace degradation in the serial recall task: When CAT helps RAT, but not MAN. *International Journal of Psychology, 34,* 301–307. doi:10.1080/002075999399602

Forde, E. M., & Humphreys, G. W. (2002). The role of semantic knowledge in short-term memory. *Neurocase, 8,* 13–27. doi:10.1093/neucas/8.1.13

Foygel, D., & Dell, G. S. (2000). Models of impaired lexical access in speech production. *Journal of Memory and Language, 43,* 182–216. doi:10.1006/jmla.2000.2716

Gregg, V. H., Freedman, C. M., & Smith, D. K. (1989). Word frequency, articulatory suppression and memory span. *British Journal of Psychology, 80,* 363–374.

Gupta, P., Lipinski, J., & Aktunc, E. (2005). Reexamining the phonological similarity effect in immediate serial recall: The roles of type of similarity, category cuing, and item recall. *Memory & Cognition, 33,* 1001–1016.

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

Hamilton, A. C., & Martin, R. C. (2007). Semantic short-term memory deficits and resolution of interference: A case for inhibition? In D. S. Gorfein & C. M. MacLeod (Eds.), *Inhibition in Cognition* (pp. 239–257). Washington, D. C.: American Psychological Association.

Hanley, J. R., & Bakopoulou, E. (2003). Irrelevant speech, articulatory suppression, and phonological similarity: A test of the phonological loop model and the feature model. *Psychonomic Bulletin & Review, 10,* 435–444.

Henson, R. N. A. (1998). Short-term memory for serial order: The Start-End Model. *Cognitive Psychology, 36,* 73–137. doi:10.1006/cogp.1998.0685

Henson, R. N. A., Norris, D. G., Page, M. P. A., & Baddeley, A. D. (1996). Unchained memory: Error patterns rule out chaining models of immediate serial recall. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 49*A, 80–115. doi:10.1080/027249896392810

Hodges, J. R., Patterson, K., Oxbury, S., & Funnell, E. (1992). Semantic dementia: Progressive fluent aphasia with temporal lobe atrophy. *Brain, 115,* 1783–1806. doi:10.1093/brain/115.6.1783

Hulme, C., Maughan, S., & Brown, G. D. (1991). Memory for familiar and unfamiliar words: Evidence for a long-term memory contribution to short-term memory span. *Journal of Memory and Language, 30,* 685–701. doi:10.1016/0749-596X(91)90032-F

Hulme, C., Roodenrys, S., Schweickert, R., Brown, G. D. A., Martin, S., & Stuart, G. (1997). Word-frequency effects on short-term memory tasks: Evidence for a redintegration process in immediate serial recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 1217–1232. doi:10.1037/0278-7393.23.5.1217

Jefferies, E., Frankish, C. R., & Lambon Ralph, M. A. (2006a). Lexical and semantic binding in verbal short-term memory. *Journal of Memory and Language, 54,* 81–98. doi:10.1016/j.jml.2005.08.001

Jefferies, E., Frankish, C., & Lambon Ralph, M. A. (2006b). Lexical and semantic influences on item and order memory in immediate serial recognition: Evidence from a novel task. *Quarterly Journal of Experimental Psychology, 59,* 949–964. doi:10.1080/02724980543000141

Jefferies, E., Frankish, C., & Noble, K. (2009). Lexical coherence in short-term memory: Strategic reconstruction or "semantic glue"? *Quarterly Journal of Experimental Psychology, 62,* 1967–1982. doi:10.1080/17470210802697672

Jefferies E., Hoffman P., Jones R. W., & Lambon Ralph M. A. (2008). The impact of semantic impairment on verbal short-term memory in stroke aphasia and semantic dementia: A comparative study. *Journal of Memory and Language, 58,* 66–87.

Jefferies, E., & Lambon Ralph, M. A. (2006). Semantic impairment in stroke aphasia versus semantic dementia: A case-series comparison. *Brain: A Journal of Neurology, 129,* 2132–2147.

Kane, M. J., & Engle, R. W. (2000). Working-memory capacity, proactive interference, and divided attention: Limits on long-term memory retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 336–358. doi:10.1037/0278-7393.26.2.336

Knott, R., Patterson, K., & Hodges, J. R. (2000). The role of speech production in auditory-verbal short-term memory: Evidence from progressive fluent aphasia. *Neuropsychologia, 38,* 125–142. doi:10.1016/S0028-3932(99)00069-X

Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22,* 1–38. doi:10.1017/S0140525X99001776

Levy, B. (1971). Role of articulation in auditory and visual short-term memory. *Journal of Verbal Learning and Verbal Behaviour, 10,* 123–132. doi:10.1016/S0022-5371(71)80003-8

Lian, A., Karlsen, P. J., & Eriksen, T. B. (2004). Opposing effects of phonological similarity on item and order memory of words and nonwords in the serial recall task. *Memory, 12,* 314–337. doi:10.1080/09658210344000426

Logie, R. H., Della Sala, S., Laiacona, M., Chalmers, P., & Wynn, V. (1996). Group aggregates and individual reliability: The case of verbal short-term memory. *Memory & Cognition, 24,* 305–321.

Martin, R. C., Lesch, M., & Bartha, M. (1999). Independence of input and output phonology in word processing and short-term memory. *Journal of Memory and Language, 41,* 3–29. doi:10.1006/jmla.1999.2637

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. doi:10.1006/jmla.1994.1005

Masson, M. E. J., & Loftus, G. R. (2003). Using confidence intervals for graphically based data interpretation. *Canadian Journal of Experimental Psychology, 57,* 203–220. doi:10.1037/h0087426

Murdock, B. B., Jr., & Vom Saal, W. (1967). Transpositions in short-term memory. *Journal of Experimental Psychology, 74,* 137–143. doi: 10.1037/h0024507

Murray, D. J. (1968). Articulation and acoustic confusability in short-term memory. *Journal of Experimental Psychology, 78,* 679–684. doi: 10.1037/h0026641

Murray, D. J., & Newman, F. M. (1973). Visual and verbal coding in short-term memory. *Journal of Experimental Psychology, 100,* 58–62. doi:10.1037/h0035506

Oberauer, K. (2002). Access to information in working memory: Exploring the focus of attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 411–421. doi:10.1037/0278-7393.28.3.411

Page, M. P. A., Madge, A., Cumming, N., & Norris, D. G. (2007). Speech errors and the phonological similarity effect in short-term memory: Evidence suggesting a common locus. *Journal of Memory and Language, 56,* 49–64. doi:10.1016/j.jml.2006.09.002

Patterson, K., Graham, N., & Hodges, J. R. (1994). The impact of semantic memory loss on phonological representations. *Journal of Cognitive Neuroscience, 6,* 57–69. doi:10.1162/jocn.1994.6.1.57

Peterson, L. R., & Peterson, M. J. (1959). Short-term retention of individual verbal items. *Journal of Experimental Psychology, 58,* 193–198. doi:10.1037/h0049234

Plaut, D. C., & Kello, C. T. (1999). The emergence of phonology from the interplay of speech comprehension and production: A distributed connectionist approach. In B. MacWhinney (Ed.), *The emergence of language* (pp. 381–415). Mahwah, NJ: Erlbaum.

Poirier, M., & Saint-Aubin, J. (1995). Memory for related and unrelated words: Further evidence on the influence of semantic factors in immediate serial recall. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 48*A, 384–404.

Poirier, M., & Saint-Aubin, J. (1999). The influence of long-term memory factors on immediate serial recall: An item and order analysis. *International Journal of Psychology, 34,* 347–352. doi:10.1080/002075999399675

Postle, B. R. (2006). Working memory as an emergent property of mind and brain. *NeuroImage, 30,* 950–962. doi:10.1016/j.neuroimage.2005.10.018

Proctor, R. W., & Fagnani, C. A. (1978). Effects of distractor-stimulus modality in the Brown-Peterson distractor task. *Journal of Experimental Psychology: Human Learning and Memory, 4,* 676–684. doi:10.1037/0278-7393.4.6.676

Romani, C., McAlpine, S., & Martin, R. C. (2008). Concreteness effects in different tasks: Implications for models of short-term memory. *The Quarterly Journal of Experimental Psychology, 61,* 292–323. doi:10.1080/17470210601147747

Ruchkin, D. S., Grafman, J., Cameron, K., & Berndt, R. S. (2003). Working memory retention systems: A state of activated long-term memory. *Behavioral and Brain Sciences, 26,* 709–728. doi:10.1017/S0140525X03000165

Saffran, E. M., & Martin, N. (1997). Language and auditory-verbal short-term memory impairments: Evidence for common underlying processes. *Cognitive Neuropsychology, 14,* 641–682. doi:10.1080/026432997381402

Saint-Aubin, J., & Poirier, M. (2000). Immediate serial recall of words and nonwords: Tests of the retrieval-based hypothesis. *Psychonomic Bulletin & Review, 7,* 332–340.

Schweickert, R. (1993). A multinomial processing tree model for degradation and redintegration in immediate recall. *Memory & Cognition, 21,* 168–175.

Shallice, T., & Warrington, E. K. (1970). Independent functioning of verbal memory stores: A neuropsychological study. *The Quarterly Journal of Experimental Psychology, 22,* 261–273. doi:10.1080/00335557043000203

Shattuck-Hufnagel, S. (1979). Speech errors as evidence for a serial order mechanism in sentence production. In W. E. Cooper & E. C. Walker (Eds.), *Sentence processing: Psycholinguistic studies presented to Merrill Garrett* (pp. 295–342). Hillsdale, NJ: Erlbaum.

Storkel, H. L. (2004). Methods for minimizing the confounding effects of word length in the analysis of phonotactic probability and neighborhood density. *Journal of Speech, Language, and Hearing Research, 47,* 1454–1468. doi:10.1044/1092-4388(2004/108)

Stuart, G., & Hulme, C. (2000). The effects of word co-occurrence on short-term memory: Associative links in long-term memory affect short-term memory performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 796–802. doi:10.1037/0278-7393.26.3.796

Treiman, R., & Danis, C. (1988). Short-term memory errors for spoken syllables are affected by the linguistic structure of the syllables. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 145–152. doi:10.1037/0278-7393.14.1.145

Vallar, G., & Baddeley, A. D. (1984). Fractionation of working memory: Neuropsychological evidence for a phonological short-term store. *Journal of Verbal Learning and Verbal Behavior, 23,* 151–161. doi:10.1016/S0022-5371(84)90104-X

Vitevitch, M. S., & Luce, P. A. (2004). A Web-based interface to calculate phonotactic probability for words and nonwords in English. *Behavior Research Methods, Instruments & Computers, 36,* 481–487.

Walker, I., & Hulme, C. (1999). Concrete words are easier to recall than abstract words: Evidence for a semantic contribution to short-term serial recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 1256–1271. doi:10.1037/0278-7393.25.5.1256

Watkins, M. J. (1977). The intricacy of memory span. *Memory & Cognition, 5,* 529–534.

Wickelgren, W. A. (1965). Acoustic similarity and intrusion errors in short-term memory. *Journal of Experimental Psychology, 70,* 102–108. doi:10.1037/h0022015

Wickens, D. D. (1973). Some characteristics of word encoding. *Memory & Cognition, 1,* 485–490.

Wong, W., & Law, S.-P. (2008). The relationship between semantic short-term memory and immediate serial recall of known and unknown words and nonwords: Data from two Chinese individuals with aphasia. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34,* 900–917. doi:10.1037/0278-7393.34.4.900