January 2019

Experimentally Examining the Proposed Relationships Among “Rehearsal-based” Effects

Corey Ian McGill
cmcgil1@lsu.edu

EXPERIMENTALLY EXAMINING THE PROPOSED RELATIONSHIPS AMONG “REHEARSAL-BASED” EFFECTS

A Dissertation

Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

in

The Department of Psychology

by
Corey Ian McGill
B.S., Louisiana State University, 2013
B.S., Louisiana State University, 2016
May 2019

Table of Contents

Abstract

Despite the importance of rehearsal to most models of verbal working memory, its role has been recently called into question. Much prior work in support of rehearsal models has centered on the experimental effects of word-length, phonological-similarity, and irrelevant sound on serial order recall performance and the interaction of all three with concurrent articulation. However, recent research has suggested that confounding effects of stimuli, such as orthographic neighborhood, may be the true cause of the word-length effect. While these findings alone have significant implications for modern models of rehearsal, to understand them within the context of modern theories of working memory, they must also be examined through the lens of the phonological-similarity and irrelevant-sound effects. Thus, through a series of three experiments, the influence of neighborhood in each of these effects was assessed, using strict controls for both orthographic and phonological neighborhood size. The word-length effect was significantly reversed; longer words were significantly better recalled than short words. However, the phonological-similarity effect remained significant even when neighborhood size was controlled. The irrelevant-sound effect was significant when stimuli had no orthographic or phonological neighbors, but was eliminated when stimuli had both. These findings present significant problems for common memory models that include a role for rehearsal, as the relationship between "rehearsal-based" effects was more tenuous than may have otherwise been anticipated.

Introduction

Very few constructs have been used more frequently in models of memory than rehearsal. Rehearsal, or the recitation of to-be-remembered items in order to prevent memory loss, is an important component for a number of models of memory, specifically, in models of verbal short-term or working memory (e.g., Baddeley, 1986; 2000; Camos, Lagner, & Barrouillet, 2009; Cowan, 2005). Working memory is an area of memory that deals with immediate perceptual and conscious processing of information as well as the storage of recently processed information. In most models of working memory, the sub-vocal recitation of to-be-remembered information counteracts the effects of memory loss caused by time-based forgetting/decay by maintaining recent information within the working memory system.

Traditionally, four experimental effects have been attributed to, and used as evidence for, the role of rehearsal within working memory: the effects of concurrent articulation, the word-length effect (WLE), the phonological-similarity effect (PSE), and the irrelevant-sound effect (ISE). Each has a long history of attribution to the process of sub-vocal articulation, the non-verbal speech of task-relevant information (Baddeley, Lewis, & Vallar, 1984; Baddeley, Thompson, & Buchanon, 1975; Miles, Jones, & Madden, 1991). However, more recent research has called into question the role of time-based forgetting, and specifically, the role of rehearsal in working memory (Farrell, Oberauer, Greaves, Pasiecznik, Lewandowsky, & Jarrold, 2016; Lewandowsky & Oberauer, 2015; Lewandowsky, Oberauer, & Brown, 2009). The current research examined the viability of the WLE, PSE, and ISE being caused by a single mechanism, and evaluated what role the sub-vocal speech of to-be-remembered items might have within a possible shared cause.

The Traditional Role of Rehearsal within Working Memory

Murray (1967) identified possibly the most influential effect in support of rehearsal within working memory, concurrent articulation. According to Google Scholar, this original article has been cited over 200 times. Murray determined that when individuals are asked to repeatedly recite task irrelevant information, performance on a number of short-term memory tasks significantly decreases. This finding has been replicated numerous times (e.g., Baddeley et al., 1984; Baddeley & Hull, 1979; Baddeley et al., 1975; Camos et al., 2009; Larsen & Baddeley, 2003; Jalbert, Neath, & Surprenant, 2011; Neath, Farley, & Suprenant, 2003). While concurrent articulation effects can be accounted for in a number of ways (e.g., interference, Nairne, 1990; 2002), the most common account is that overt verbal speech limits the ability for individuals to sub-vocally recite (rehearse) the to-be-remembered information, which is thought to help overcome the problems of a capacity limited system (Baddeley, 1986; Camos et al., 2009; Cowan 2005).

Limiting rehearsal through concurrent articulation has a clear impact on memory performance, making it an important hallmark for time-based forgetting. Theories including a role for time-based forgetting have almost exclusively proposed that limiting the rehearsal process with concurrent articulation limits, or even prevents, individuals from bringing recently presented information back into the forefront of memory to prevent the decay of that information. Repeated rehearsal of information lessens the amount of time between the last instance of the to-be-remembered items in memory and the current moment, thus increasing the likelihood of correct recall. However, when the rehearsal process is limited, the information decays because the traces cannot be refreshed or revived. While other accounts of concurrent articulation can just as adequately explain its effects (see below), without the demonstrable effects of articulation, it would be very difficult to conclude that rehearsal is used to combat time-based forgetting.

Additionally, the WLE and the PSE have been believed to demonstrate the benefits of using rehearsal to maintain items in memory. The WLE is the tendency for memory span performance to be less for words of longer duration (e.g. individuals can remember more one-syllable words, like *harm*, than five-syllable words, like *organization*; Baddeley et al., 1975). Traditional accounts of the WLE suggest that if the spoken duration of an item is increased, then the amount of time that it takes to rehearse said item is increased as well. This increased time to sub-vocally articulate results in more item decay before the item(s) can be rehearsed (Baddeley, Chincotta, Stafford, & Turk, 2002; Baddeley et al., 1975; Cowan, Nugent, & Elliott, 2000). Such a finding indicates not only that efficient rehearsal facilitates the maintenance of items in working memory, but also that the temporal component of the rehearsal process is important. The relationship between the WLE and pronunciation duration is believed to indicate the link between the rehearsal process and time-based decay. The ability to rehearse more items in less time is believed to minimize the effects of decay; therefore, short items facilitate more effective rehearsal (Baddeley et al., 2002; Baddeley et al., 1975; Cowan et al., 2000).

Moreover, the PSE suggests efficient rehearsal also improves memory for item order. The PSE is the lessened memory span performance when to-be-remembered items share phonemes when spoken aloud (e.g. BCDGPTV), as opposed to when items do not share phonemes (e.g. FLMNSXZ; Baddeley et al., 1984; Murray, 1967). Working memory theories including a role for time-based forgetting explain the PSE as occurring due to interference within the rehearsal process from phonologically-similar items (Nairne, 1990). Interference is caused by the ambiguity inherent in the production of sub-vocal speech involved in the rehearsal of phonologically-similar items. The lack of distinctiveness in the speech sounds across items results in confusion of the order of the to-be-remembered information (Lian, Karlsen, & Eriksen, 2004; Spurgeon, Ward, & Matthews, 2014). The confusion of order information leads to an increased number of transposition errors (errors in which two items in the list are switched causing neither to be correctly recalled). The rehearsal process is thought to maintain order information in much the same way it is thought to maintain item information (Beaman & Jones, 1997; Larsen & Baddeley, 2003). The repeated rehearsal of the items in the order they were presented limits the decay of the order information and items are more likely to be recalled in the correct order. However, when items share phonemes it becomes more difficult to maintain the item order and performance suffers.

Additionally, the role of order in the PSE can be examined by comparing both free-recall and serial-order recall scoring methods on the same set of responses. A free-recall scoring system removes the importance of order maintenance from the response by scoring items as correct even if placed in the wrong order during recall. In contrast, the more standard serial-order recall scoring system only scores responses correct when items are correctly ordered at recall. By comparing the two scoring methods, the importance of order within an effect can be examined, because transposition errors will result in a response being recorded as incorrect only in serial-order recall. While the PSE has been shown to persist with free-recall scoring, the effect size is lessened (Coltheart, 1993; Coltheart & Langdon; 1998, Spurgeon et al., 2008), which has two implications. First, the PSE can result in item errors independent of the loss of order information due to the persistence of the effect. Second, the PSE does cause an increased number of transposition errors as demonstrated by the lessened effect size when those errors are no longer scored as incorrect (i.e. free-recall scoring). Taken together, this means that the PSE may be the result of two simultaneous effects stemming from the lessened rehearsal efficacy caused by the shared phonemes across stimuli. Of note, when the WLE is scored using both free and serial recall conditions, the size of the effect was not significantly changed (Coltheart & Langdon, 1998). This means that ineffective order maintenance is likely not a cause of the WLE, and errors are cause wholly by the loss of item information.

The PSE further implicates the importance of verbal information in rehearsal. While the WLE is, among other causes, believed to be related to the spoken duration, the PSE is caused by the spoken sound of the verbal name for the target items and persists even with visual presentation of the to-be-remembered items (Baddeley et al., 1984; Salamé & Baddeley, 1982). As no verbal information for the to-be-remembered items occurs automatically with visual presentation, individuals must be creating a verbal representation of the to-be-remembered items, possibly through the sub-vocal speech required of rehearsal.

Supplementary support for the significance of order maintenance in the rehearsal process comes from the ISE. The ISE is the lessened ability to perform serial order recall in the presence of changing-state auditory stimuli compared to silence (Colle & Welsh, 1976). For example, when asked to recall a list of digits in order, individuals perform worse when listening to a simultaneously presented list of irrelevant items (e.g. random letters), compared to when they are not asked to listen to any additional stimuli. Order processing is vital to the ISE, as indicated by two lines of evidence, (1) the elimination of the effect with certain methodologies that remove the importance of order maintenance (missing-item tasks; Beaman & Jones, 1997; Elliott et al., 2016) and (2) the reduction of the effect with certain auditory stimuli that do not require order processing (repetition of the same sound known as “steady-state” sounds; Jones, Macken, & Murray, 1993; Lange, 2005).

First, to remove the importance of order information, missing-item tasks require individuals to efficiently store only the presented items, but not the order of presentation. For example, participants may be visually-presented with six digits with random selection without replacement from the numbers one to seven, and in a missing-item task participants need only identify which digit was not presented to respond (Buschke, 1963). Similarly, a probed-recall task may present the same six digits with the same selection criteria, but require participants to indicate which digit was presented after another. In both tasks only the response criteria changes, but both show differential effects of changing-state auditory stimuli, or stimuli that differ from item to item. Performance in the missing-item task is not susceptible to the ISE, but performance in the probed-recall task is (e.g., Beaman & Jones, 1997; Elliott et al., 2016). It is believed that the different response criteria across the two types of tasks either emphasizes or eliminates the importance of order in the task. In a probed-recall trial, participants must maintain not only which items were presented, but the order of presentation as well. In contrast, during a missing-item trial, participants need only to maintain which items were presented, but not the order they were presented. Thus, if the ISE can be eliminated by removing the importance of order in the task, then changing-state auditory stimuli (e.g. different letters presented in succession) must specifically interfere with the order processing/maintenance of to-be-remembered items (Beaman & Jones, 1997; Elliott et al., 2016).

Second, elimination of order information in the irrelevant auditory channel eliminates or significantly reduces the ISE. For example, unlike changing-state auditory stimuli, steady-state auditory stimuli (e.g. the same letter repeated) produce little to no significant effects on serial order recall (Elliott et al., 2016; Jones et al., 1993; Lange, 2005). This is because no information is gained by remembering the order of the repetitive stimulus in the irrelevant auditory channel (Hughes, Vachon, & Jones, 2007; Schröger, 1997). For example, if you are presented with the same auditory stimulus repeatedly then there is no obligatory order processing because no information is gained by trying to order identical information. Therefore, it is hypothesized by some researchers that the ISE is the result of automatic order processing of auditory stimuli which interrupts the maintenance of order information in the rehearsal process (Hughes, 2014; Hughes, Chamberland, Tremblay, & Jones, 2016).

In addition to specific reasons for the WLE, PSE and ISE being attributed to rehearsal, all three phenomena interact with concurrent articulation. This suggests that they may all share a similar cause, and that this cause is limited by articulation of task irrelevant information (Baddeley et al., 1975; Murray, 1967; McGill & Elliott, in prep). For each of the above effects, when concurrent articulation is required, they are eliminated or at least significantly reduced (Baddeley et al., 1975; Hanley, 1997; Murray, 1967). Such an interaction has been considered a requirement for rehearsal effects, because it is believed that individuals cannot sub-vocally recite to-be-remembered information while verbally reciting other information. Furthermore, an interaction with concurrent articulation is viewed as strong support for the role of rehearsal in causing an effect, because the elimination of sub-vocal speech is the most obvious process that concurrent articulation limits. For example, other than order importance in the ISE and some indirect evidence (e.g. ruling out other explanations like perceptual effects through the temporal separation of TBR item presentation and the presentation of irrelevant sound; Macken, Mosdell, & Jones, 1999; Miles et al., 1991), there is little to no direct evidence suggesting rehearsal as a cause beyond the elimination of the ISE under concurrent articulation.

Issues with the Traditional Model of Rehearsal

While elegant to posit a singular mechanism, the role of rehearsal in the aforementioned articulatory effects needs to be interpreted with caution. Without proper examination, attribution of all effects that interact with concurrent articulation to rehearsal may lead to erroneous assumptions about the role of rehearsal in working memory. For example, when rehearsal is limited through other means, such as speeded presentation, the WLE (Coltheart & Langdon, 1998; McGill & Elliott, in prep), the PSE (Coltheart & Langdon, 1998; McGill & Elliott, in prep), and the ISE (McGill & Elliott, in prep) have been shown to persist. More specifically, McGill and Elliott (in prep) found no reduction in the size of the WLE or ISE when to-be-remembered items were presented at a rate of four items/s, while the PSE was significantly reduced, as compared to a presentation rate of one item/s. However, the effect was not eliminated. This finding suggests that while speeded presentation does interact with the PSE, it has no impact on the size of the WLE or ISE. This set of findings presents two potential problems for the traditional model of rehearsal. First, if all three effects share a single cause (sub-vocal speech), then experimental manipulations interacting with one effect should similarly interact with the others. Second, it would be expected that reducing the amount of time individuals must rehearse items would result in a significant reduction in the size of any rehearsal effects under a traditional model of rehearsal. The rapid presentation rate should not allow individuals time to sub-vocally recite much more than the current item presented on screen. That inability to effectively rehearse previous items, in addition to the immediate recall prompt, should combine to both limit the amount of decay that occurs and individuals' ability to combat any decay that might occur. The immediate recall prompt further limited rehearsal by allowing participants to respond immediately after the presentation of the final to-be-remembered item, as opposed to including a retention interval between the final-item presentation and when participants are allowed to respond. Therefore, the persistent and equivalent effect in speeded presentation suggests that if sub-vocal speech is the cause of both the WLE and the ISE, a single sub-vocal utterance at the presentation of each item is enough to cause the effect, and that the use of cumulative rehearsal to limit time-based decay may not be the cause of either effect. It is also possible that speeded presentation results in individuals using a unique strategy to recall items, but such an assertion requires that strategy to result in word-length and irrelevant sound effects that do not significantly differ from the traditional effects.

Further problems for the traditional role of rehearsal in the WLE are highlighted by the unreliability of multiple methodological manipulations within the WLE that seem to be impacted by stimulus selection. For example, many attempts to directly replicate the WLE with words matched for phonemic complexity across the original lists used in Baddeley et al. (1975) have been successful (Bireta, Neath, & Surprenant, 2003; Cowan et al., 1992; Longoni, Richardson, & Aiello, 1993; Lovatt, Avons, & Masterson, 2000; Nairne, Neath, & Serra, 1997). However, these results have not been replicated using different stimuli that were also matched for phonemic complexity across pronunciation duration (Bireta et al, 2003; Caplan, Rochon, & Waters, 1992; Lovatt et al., 2000; Service, 1998). This ability to demonstrate a WLE that persists with controlled phonemic complexity is vital to the traditional rehearsal explanation. If significant differences in pronunciation duration cannot produce significant effects without the increased phonemic complexity often confounded with longer words, it becomes impossible to conclude the time between rehearsal utterances in longer words causes significantly more decay. Instead, it can be argued that the increased complexity of longer words causes increased inter-item interference (Lewandowsky & Farrell, 2000; Lewandowsky & Oberauer, 2015; Lewandowsky et al., 2009; Nairne, 1990; 2002; Neath, 2000). Additionally, contradictory results have been observed within the WLE when the same list includes both short and long items (e.g., Cowan, Baddeley, Elliott, & Norris, 2003; Hulme, Suprenant, Bireta, Stuart, & Neath, 2004), suggesting that stimulus selection may contribute to some of the contradictory results in the WLE. While the intricacies of the prior papers are not presently reviewed, of importance is that when Bireta, Neath, and Surprenant (2006) varied the stimuli and methodology from the conflicting reports of Cowan et al. (2003) and Hulme et al. (2004), it was found that the stimuli used in each experiment could entirely account for the differential results. That is, when the methodologies from one experiment were used with the stimuli from another, the results replicated those of the experiment from which the *stimuli* were selected, not the methodology.

Effects of Neighborhood Size

Jalbert, Neath, Bireta, and Surprenant (2011) suggested some important variable that was not commonly controlled for may account for much of the seemingly contradictory effects within prior WLE work. While many variables are controlled for across short- and long-word items in prior work (e.g. phonemic complexity, familiarity, frequency, etc.), it was proposed that the number of orthographic neighbors that to-be-remembered words have impacts the WLE and may even account for the differential findings in prior work. Orthographic neighbors are the number of words that can be formed by changing a single letter in the to-be-remembered word (e.g. *cat* has orthographic neighbors including *bat*, *cot*, and *cab*). Prior work had established that orthographic neighborhood size affects the lexical access of both words and non-words (Andrews, 1989; Sears, Hino, & Lupker, 1995), suggesting that orthographic neighbors may facilitate individuals' ability to correctly access words at recall. This is hypothesized to occur because to-be-remembered words at least partially activate orthographically similar words, and that this pattern of activation can be used to facilitate the correct recall of the to-be-remembered items. Jalbert, Neath, Bireta, and Surprenant (2011) demonstrated that, generally and in prior work on the WLE, shorter words tend to have more orthographic neighbors than longer words, even when other variables are controlled. They further hypothesized that orthographic neighborhood size may be contributing to both the WLE as well as the difficulty in replicating some findings depending on stimulus selection. It was found that orthographic neighborhood size effects significantly impacted memory, even when phonemic complexity was controlled, as lists of words with a larger orthographic neighborhood were recalled better and faster than words with a smaller orthographic neighborhood, for both serial reconstruction of order and verbal recall. Finally, when orthographic neighborhood size was controlled for, there was no significant difference between one- and three-syllable words on memory performance. It was suggested that the larger orthographic neighborhood typical of shorter words is the true cause of the WLE, and not the duration of the word. Such a finding is difficult to account for under a rehearsal explanation of the WLE alone. In order for rehearsal to play a role in the WLE, it must also coexist with a neighborhood size effect occurring simultaneously.

In a follow-up examination, Jalbert, Neath, and Surprenant (2011) found that the orthographic neighborhood effect could be eliminated by articulatory suppression, applied to non-word stimuli, and fully crossed with word-length. When fully crossed, neighborhood size significantly impacted performance while no significant main effect of word-length or interaction between word length and neighborhood size were found. If traditional rehearsal were, in part, responsible for the WLE, then the syllabic-length should result in a significant effect unique from neighborhood size in addition to any effects of neighborhood size. The lack of such a finding was interpreted to be potentially devastating to the rehearsal explanation of the WLE.

However, Guitard, Saint-Aubin, Tehan, and Tolan (2017) further explored these recent findings by additionally controlling for the number of unigrams, bigrams, and trigrams within to-be-remembered stimuli. N-gram measures break a word down into its constituent letter combinations and examine the frequency with which those combinations appear in other words of the same length. For example, the word *picnic* contains six unigrams (*p-i-c-n-i-c*), five bigrams (*pi-ic-cn-ni-ic*), and four trigrams (*pic-icn-cni-nic*). The more frequent a word's n-grams are, the more familiar the word structure can be assumed to be (Freeman, Heathcote, Chalmers, & Hockley, 2010; Rice & Robinson, 1975), which may facilitate accurate recall. The authors proposed that the prior work controlling for orthographic neighborhood in the WLE (Jalbert, Neath, Bireta, & Surprenant, 2011; Jalbert, Neath, & Surprenant, 2011) did not adequately control for the n-gram frequency across stimuli because shorter words tended to use less common structure in order to match the size of the orthographic neighborhood of longer words. Thus, both one- and three-syllable French words were matched for both orthographic neighborhood size as well as unigram, bigram, and trigram frequency. It was found that when orthographic neighborhood size and n-gram frequency were both controlled for, there was a significant difference between both one- and three-syllable words in which one-syllable words were significantly better recalled. However, despite the compelling findings of Guitard et al. (2017), the prior volatile nature of effects due to stimuli means that additional replications with different stimuli must be performed, including stimuli in languages other than French.

Summary and Overview of the Current Experiments

The following experiments examined the possibility that these potentially confounding effects of neighborhood size in the WLE also similarly affect the PSE and ISE. First, the methodology of Guitard et al. (2017) was replicated with English stimuli in which neighborhood size and word structure were controlled for across both short and long words. The results were expected to replicate those of Guitard et al. (2017) and demonstrate a significant WLE. Such a finding would support the syllabic-length account of the WLE. Additionally, two follow-up studies examined how the PSE and ISE might be impacted by controlling for neighborhood effects. If the three effects share a similar cause, as has been often suggested, it was expected they would demonstrate similar patterns of results regarding the effects of neighborhood size. If that shared cause is unrelated to neighborhood effects, then all three effects should be demonstrated when neighborhood size is controlled. However, if all three effects are related, and demonstrate an effect of neighborhood size, it may be that the shared cause of all three is in some way related to orthographic and/or phonological neighborhood. The results of all three experiments are then discussed regarding the possibility of a shared cause based around both their shared and unique interactions with neighborhood.

Experiment 1: The Word-length Effect

Experiment 1 attempted to replicate the findings of Guitard et al. (2017) using English stimuli that were similarly matched for orthographic neighborhood and n-gram frequency as well as phonological neighborhood size. Similar to orthographic neighborhood, a word's phonological neighborhood is made up of all words that can be created by replacing a single phoneme with another. For example, the word *ghost* has phonological neighbors that include *most*, *toast*, *gust*, and *guest*. A large phonological neighborhood has been shown to have distinct effects on certain cognitive tasks. Luce and Pisoni (1998) demonstrated that words with a larger phonological neighborhood are more difficult to recognize when presented auditorily with noise, and Roodenrys, Hulme, Lethbridge, Hinton, and Nimmo (2002) found that serial order recall was improved by words with a large phonological neighborhood. While Guitard et al. (2017) did not explicitly control for phonological neighborhood, a review of the chosen stimuli using the French Cross-Linguistic Easy-Access Resource for Phonological and Orthographic Neighborhood Densities (CLEARPOND; Marian, Bartolotti, Chabal, & Shook, 2012) found no French phonological neighbors for any stimuli used in Experiments 5 and 6 (neighborhood information was not found for grief, cheptel, or cardiogramme). The lack of explicit control for phonological neighbors may be an important limitation of the stimuli used in their Experiments 5 and 6, and is an additional factor that will be addressed in the current research. Additionally, the current Experiment 1 included concurrent articulation to examine if any possible effect of word-length with items matched for neighborhood size and bigram frequency could be eliminated through limiting participants' ability to sub-vocally articulate. It was expected that words with more syllables would take longer to produce and be recalled worse than words with fewer syllables. This is a vital first step, as much of the prior work in the WLE leading to Guitard et al. (2017) was inspired by the difficulty in replicating effects with additional stimuli. Thus, we looked to validate the findings of Guitard et al. (2017) by explicitly controlling for phonological neighborhood in addition to orthographic neighborhood and bigram frequency in a set of English stimuli. Finally, it was expected that any significant WLE would be eliminated when participants are required to concurrently articulate.

Methods

Participants

Fifty-two Louisiana State University undergraduates aged 18-24 (M_{age} = 19.98, $std._{age}$ = 1.50) participated for course credit. Of the 52 participants 31 were female and 21 were male. All participants reported English as their native language, having either normal or corrected vision, and not suffering from any loss of hearing.

Design and power analysis

Experiment 1 employed an entirely within-subjects design with two independent variables, both with two levels. The first independent variable, word-length, was be manipulated by using stimuli that are either two- or four-syllables in length, and the second, articulation condition, will be manipulated by completing trials in silence, or requiring silent concurrent articulation throughout item presentation. The dependent variables were the proportion correct scores using both strict serial-position scoring and free-recall, as discussed below.

Using G-Power, a total sample size of 12 was suggested to replicate the results of Guitard et al. (2017), based on a partial eta squared of .17 for the interaction and assumed power of 0.80 (power analysis indicated power > 0.99). However, Anderson, Kelly, and Maxwell (2017) suggested the need to correct for publication bias and uncertainty when calculating power for previously published work. When both were controlled for, the Bias and Uncertainty Corrected Sample Size (BUCSS) power analysis indicated that while the interaction between word-length and articulation in Guitard et al. (2017) may not be accurately estimated, the effect of word-length alone would be replicable with a sample size of 48 while controlling for both publication bias and uncertainty.

Materials

Word-length was manipulated using two-syllable (short) or four-syllable (long) words. Two- and four-syllable words were chosen to ensure that the words were matched for word frequency, orthographic and phonological neighborhood size, unigram average, bigram average, trigram average, bigram frequency by position, picture naming response time, and concreteness (See *Table 1*). Words were not pairwise matched across long and short stimuli due to the inherent difficulty in identifying four- and two-syllable words that are similar across all controlled variables. Using basic *t*-tests to identify differences across lists allows for slightly more variability which then allowed for more variables to be explicitly controlled. However, it is important to note that in doing so, singular outliers on a controlled variable may have undue influence. Additionally, potential words were examined for possible additional neighbors or otherwise incorrect information (e.g. *cheetah* was removed as a potential two-syllable TBR item because of possible phonological neighbors *vita* and *Rita*).

Table 1. Long (four-syllable) and short (two-syllable) stimuli in Experiment 1.

Word	Length	Freq_HAL	SUBTLWF	OrthoN	PhonoN	UG_Mean
automobile	10	3053	5.71	0	0	27498.02
meteoroid	9	9	n/a	0	0	33107.79
brontosaurus	12	50	0.22	0	0	26494.94
formaldehyde	12	279	0.67	0	0	25471.44
terracotta	10	40	0.1	0	0	34691.82
ukulele	7	64	0.57	0	0	24231.31
kaleidoscope	12	173	0.29	0	0	27647.96
videotape	9	2152	5.18	0	0	30848.27
geologist	9	817	1	0	0	28384.03
elevator	8	3215	24.41	0	0	33215.29
M_{Long}	**9.8**	**985.2**	**4.24**			**29159.09**
p-value	0	0.792	0.496			0.132
M_{Short}	**6.9**	**853.5**	**9.58**			**26531.78**
nostril	7	376	0.69	0	0	30100.29
picnic	6	1374	11.69	0	0	22029.17
debris	6	1761	3.12	0	0	27672.58
trapeze	7	152	1.35	0	0	31365.93
cauldron	8	1019	0.47	0	0	23755.62
lozenge	7	57	0.16	0	0	28268.77
musket	6	191	0.98	0	0	24625.94
thermos	7	306	1.12	0	0	30646.93
burglar	7	659	5.53	0	0	19868.72
upstairs	8	2640	70.73	0	0	26983.89

Note. All data from the English Lexicon Project (Balota et al., 2007) or the Leipzig Corpora Collection (Goldhahn et al., 2012); Freq_HAL = word frequency reported by the HAL study; SUBTLWF = word frequency from Brysbaert & New (2009); OrthoN = number of orthographic neighbers; PhonoN = number of phonological neighbors; UG_Mean = average unigram frequency for all unigrams within a word

table cont'd.

Word	BG_Mean	BGFreqPOS	TG_Mean	Syllables	I_Mean_RT	Concreteness
automobile	1197.78	1114	108.29	4	679.87	4.96
meteoroid	1728.13	1967	122.5	4	825.27	4.46
brontosaurus	1567.55	1971	225.24	4	926	4.52
formaldehyde	1330.27	1994	244.29	4	923.6	4.61
terracotta	2404	1979	269.11	4	981.26	4.29
ukulele	1607.67	1290	90.96	4	891.17	4.62
kaleidoscope	1358.82	1222	109.8	4	845.86	4.79
videotape	1223	1228	148.05	4	677.24	4.92
geologist	1453.5	1217	175.32	4	707.59	4.41
elevator	2056.86	1338	225	4	642.63	4.79
M_{Long}	**1592.76**	**1532**	**171.86**		**810.05**	**4.64**
p-value	0.635	0.961	0.427		0.691	0.911
M_{Short}	**1514.05**	**1524**	**321.32**		**784.43**	**4.63**
nostril	1867.17	2131	260.46	2	705.1	4.89
picnic	1385.6	1072	52.29	2	677.48	4.83
debris	1809.6	1797	138.49	2	688.79	4.69
trapeze	1313	1843	137.84	2	833.23	4.55
cauldron	1540	1620	152.22	2	794.08	4.61
lozenge	1825.33	1328	99.26	2	1166.92	4.59
musket	860.4	1175	86.31	2	852.87	4.67
thermos	1937.67	1333	1954.7	2	825.31	4.67
burglar	1266.33	1449	89.88	2	713.41	4.44
upstairs	1335.43	1492	241.81	2	587.15	4.33

Note. BG_Mean = average bigram frequency for all bigrams within a word; BGFreqPOS = sum of the bigram frequency in the same position; TG_Mean = average trigram frequency for all trigrams within a word; I_Mean_RT = mean response time on a lexical decision task.

Procedure

Before beginning the experiment, all participants were read all potential TBR items out loud to ensure they knew how to pronounce the words. Any questions about the meaning of an item were answered.

Participants completed two blocks of experimental trials. Each block consisted of four practice trials and 40 critical trials (20 two-syllable and 20 trials of four-syllable lists), randomly ordered with the condition that no stimulus condition be repeated more than twice in a row.

Blocks were either entirely concurrent articulation or in silence, and were counterbalanced across participants. Each trial began when participants initiated the trial by pressing space. Similar to Jalbert, Neath, and Surprenant (2011), on each trial six of the ten possible items were selected in random order without replacement, and presented visually one at a time for 1000 ms each. After the presentation of the sixth item, participants were immediately presented all ten possible items in alphabetical order and participants asked to click on the six items in the order they were presented. After six items had been selected, the trial ended and participants were prompted to start the next trial.

During trial blocks in which concurrent articulation was required, participants were required to repeatedly silently recite "*one, two*" throughout the presentation of to-be-remembered items. Silent concurrent articulation was used to ensure that the articulation condition did not introduce additional auditory distraction effects while still limiting participants' ability to sub-vocally recite the to-be-remembered items. Additionally, after their response had been recorded on each trial they were prompted with an additional screen asking if they remembered to silently articulate throughout presentation. Any trials in which a participant indicated that they forgot to articulate throughout presentation were excluded from analysis, and any individuals who indicated they forgot to articulate on four or more trials within a single block were entirely excluded and replaced with another participant.

Despite being prompted to click the item in correct serial order, performance was scored with both strict serial position scoring and free recall scoring. This allowed for independent analysis of both item and order maintenance.

After the end of the experimental trials, all participants were recorded reading two lists of stimuli. The first list was a randomized list of either all ten two-syllable or four syllable words, and the second was the other list. Participants were instructed to read each list one at a time as fast as possible with the experimenter providing at least one spoken example. Once the participant understood the instructions, they read each list one at a time and were recorded through a microphone.

Results

To first assess that the pronunciation duration of the short and long stimuli matched the increased syllabic length, overall time to recite the lists was compared across the two groups. Participants recited the randomized 10-item list of two-syllable words (M = 5.98 s SD = 1.93 s) faster than the randomized 10-item list of four-syllable words (M = 6.86 s $SD.$ = 1.82 s), $t(54) = -3.61$, $p < 0.001$, which indicated that the long stimuli not only had more syllables but took longer to pronounce.

A visual representation of the WLE in Experiment 1, both with and without silent articulation for both serial-order and free recall scoring can be found in Figure 1. As there were no effects of the counter-balanced block order in either serial- or free-recall conditions, all analysis presented were collapsed across block order. The first two-way within subjects ANOVA indicated significant main effects of both word-length (long words remembered better than short words), $F(1,47) = 31.51$, $p < 0.01$, $\eta^2_{partial} = .40$, and articulation condition (performance in silence better than under silent articulation), $F(1,47) = 57.74$, $p < 0.01$, $\eta^2_{partial} = .55$, in serial-order scoring. However, there was no significant interaction between word-length and

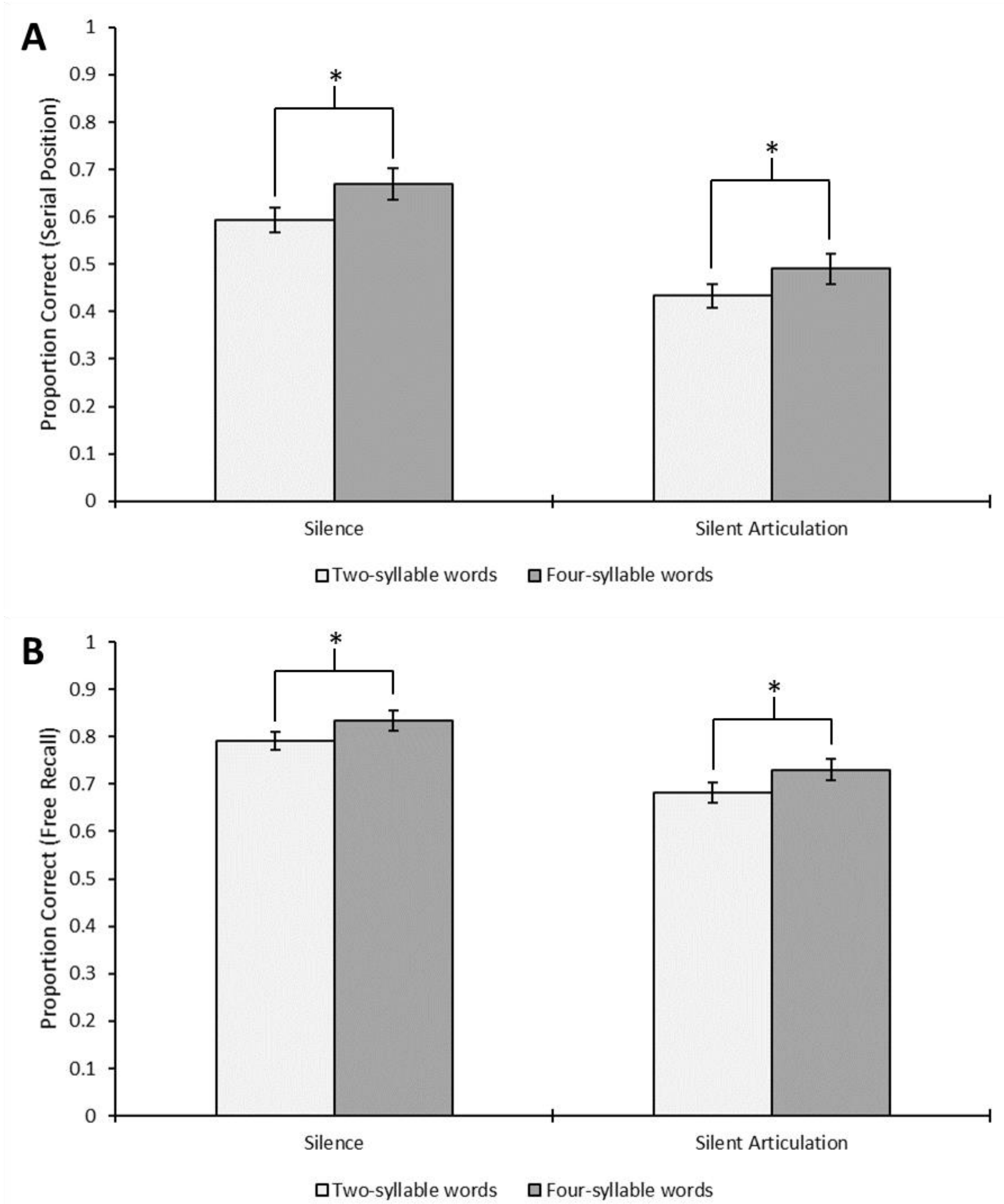

Figure 1. Proportion correct of short (two-syllable) and long (four-syllable) words under silence and silent concurrent articulation with serial-recall scoring (A) and with free recall scoring (B) in Experiment 1.

articulation condition, $F(1,47) = 0.93$, $p = 0.34$, $\eta^2_{partial} = .02$. These results indicated a reversed word-length effect that persisted under silent concurrent articulation.

The second, free-recall, two-way ANOVA replicated the findings of the first, serial-order recall, ANOVA. There were significant main effects of word-length (long words > short words), $F(1,47) = 50.01$, $p < 0.01$, $\eta^2_{partial} = .52$, and articulation condition (silence > silent articulation), $F(1,47) = 49.81$, $p < 0.01$, $\eta^2_{partial} = .52$. Additionally, there was again no significant interaction between word-length and articulation condition, $F(1,47) = 0.15$, $p = 0.70$, $\eta^2_{partial} < .01$, which again indicated a reversed WLE where long words were recalled significantly better than short words. That pattern occurred in the silent concurrent articulation as well as in silence.

Discussion

When orthographic and phonological neighborhood, word frequency, unigram average, bigram average, trigram average, bigram frequency by position, picture naming response time, and concreteness were controlled, and syllabic length, a reversed WLE was found where recall was improved for lists of four-syllable words than lists of two-syllable words. This finding appeared to indicate that within much of the prior WLE literature, two opposing effects were occurring simultaneously, an increased ability to reconstruct a memory trace from the increased information in longer words, and the worsened recall caused by a lessened orthographic and/or phonological neighborhood (Jalbert, Neath, & Surprenant, 2011). Additionally, the beneficial effects of word-length were not affected by silent concurrent articulation and persisted in free-recall scoring. Thus, the improvements in recall from increased word-length were due to improved memory for item information and not only improving the recall of order information.

Furthermore, those improvements were not reliant upon the sub-vocal articulation of the to-be-remembered items.

While the results of Experiment 1 did not replicate the findings of Guitard et al. (2017) as hypothesized, the original findings of the effects of controlled neighborhood size in the WLE were replicated (Jalbert, Neath, Bireta, & Suprenant, 2009; Jalbert, Neath, & Suprenant, 2009). However, there are a few differences in stimulus selection between the current Experiment 1 and Guitard et al. (2017). First, while both experiments did control for word complexity through n-gram information, word frequency, and orthographic neighborhood, Experiment 1 additionally used explicit controls for phonological neighborhood, concreteness, and response times in a lexical decision task. Second, Guitard et al. (2017) used additional unigram and trigram frequency by position controls similar to the bigram frequency by position of Experiment 1. Contradictory results when using similar controls during stimulus selection is not a new phenomenon in the WLE literature (Bireta et al., 2006; Cowan et al., 2003; Hulme et al., 2004), but the number of explicit controls in Experiment 1 makes identification of additional English stimuli difficult. Thus, additional replication in other languages is needed, and could be important for fleshing out what may have caused the difference in results from Guitard et al. (2017) to the current Experiment 1.

However, even if additional English stimuli cannot be used to replicate these findings in the WLE, there remains the possibility that similar controls can be used in selecting phonologically-similar words to examine the relationship between the two effects. As discussed above, the PSE has often been thought of as evidence for the importance of rehearsal in a manner similar to the WLE. However, while the WLE literature has significantly evolved in recent years, the same cannot be said of the PSE literature. This is likely due to the fact that the PSE can be explained as the result of increased inter-item interference caused by words that share phonemes (Nairne 1990; 2002; Neath, 2000). The shared phonemes across items lessen the unique features of each item that are used to correctly recall to-be-remembered items at recall, and rehearsal is not necessarily required for such an outcome to occur. For example, if all words in a list start with a unique phoneme, the initial phoneme alone could be used as a cue unique to a single item in the list, but when all words start with the same phoneme, it would eliminate that potentially facilitating cue.

While both rehearsal and interference accounts of the PSE can explain the base effect well, interference theories have difficulty explaining the interaction between the PSE and concurrent articulation when compared to rehearsal theories. As concurrent articulation eliminates the sub-vocal recitation of items that is required for rehearsal, the rehearsal explanation of the interaction is quite simple. When concurrent articulation is required, phonologically-dissimilar items no longer benefit unequally from the rehearsal process (Baddeley et al., 1984). The null effects of rehearsal are equated across both similar and dissimilar items, and no PSE occurs. However, interference accounts of the PSE suggest that concurrent articulation introduces similarity across all items regardless of their phonological-similarity (Nairne, 1990). By requiring an individual to recite a single word repeatedly throughout the presentation of the to-be-remembered items, the repeated word is encoded with the to-be-remembered item and then introduces a level of similarity across to-be-remembered items which causes interference, even when the to-be-remembered stimuli are phonologically dissimilar.

When Larsen and Baddeley (2003) experimentally altered the articulation to vary at the same rate as the presentation of the to-be-remembered items, the PSE was still eliminated.

Under the interference hypothesis of the interaction between the PSE and articulation, the differential articulation conditions should have lessened the amount of similarity introduced through articulation and allowed for the PSE to occur. It was found that neither a syncopated rhythm of a single spoken word nor different spoken words resulted in a significant PSE. In fact, the magnitude of the effect reversed for both syncopated and multiple-item articulation where the phonologically-similar words were remembered *better* than the phonologically dissimilar words. This surprising finding may suggest that more demanding concurrent articulation requirements facilitated a recall strategy in which participants reconstructed item information at recall.

While it is still viable that increased inter-item interference causes the PSE, the interaction with concurrent articulation needs to be further addressed. As even the orthographic neighborhood effect in Jalbert, Neath, and Surprenant (2011), was eliminated by concurrent articulation, it is possible that even a neighborhood-based effect could be related to the PSE. If the beneficial effects of neighborhood size are eliminated through concurrent articulation, it is possible that the elimination of the PSE under the same conditions occurs for a similar reason. The PSE may be caused by a decrease in the efficacy of a phonological neighborhood when other items in the same list also contain similar phonemes. When words share more phonemes the probability that they share phonological or orthographic neighbors increases, and if that occurs their neighborhood would no longer be as effective in helping to reconstruct an item at recall. Similarly, when all items in a list are phonologically-distinct it may be that the benefits of neighborhood size can occur normally, as items are less likely to share neighbors with words made up of a more varied number of phonemes. If this were the case it could be expected that controlling for orthographic and phonological neighborhood in the PSE might have significant implications for models of rehearsal.

Experiment 2: The Phonological-similarity Effect

Experiment 2 replicated Experiment 1 with two-syllable phonologically-similar words instead of four-syllable phonologically-distinct words. It was expected that the results of Experiment 2 would replicate those of Experiment 1, in which the effect was eliminated or reversed when orthographic and phonologic neighborhood were controlled, if the WLE and PSE share a similar cause.

Methods

Participants

Fifty-three Louisiana State University undergraduates aged 18-31 (M_{age} = 20.13, SD_{age} = 2.61) participated for course credit. Once participant did not report their age. Of the 53 participants, 41 were female and 12 were male. All participants reported English as their native language, having either normal or corrected vision, and not suffering from any loss of hearing.

Design and power analysis

Experiment 2 employed a similar within-subjects design with two independent variables, both with two levels. Phonological similarity was manipulated by using stimuli that started with the letter "s" and the phoneme "/s/", and the articulation condition was the same as Experiment 1. The dependent variables were the proportion correct scores using both strict serial-position scoring and free-recall, like Experiment 1.

Power analysis using BUCSS (Anderson et al., 2017) to control for publication bias and uncertainty suggested a sample size of 8 would be needed to replicate the PSE findings of Experiment 1 from Larsen and Baddeley (2003). However, as no prior work has attempted to control for neighborhood effects within the PSE, a similar N to Experiment 1 was used to ensure enough power if the PSE was significantly reduced when controlling for neighborhood size.

Materials

Phonological-similarity was manipulated using words that either vary in the first phoneme or all start with the same first letter (s) and phoneme (/s/) in order to ensure all other critical variables could be adequately controlled. The phonologically-distinct items were the same stimuli as the two-syllable items in Experiment 1. The ten additional phonologically similar items were also two-syllables and were again matched for word frequency, orthographic and phonological neighborhood size, bigram average, bigram frequency by position, picture naming response time, and concreteness (See *Table 2*). Again, pairwise matching was not performed and potential stimuli were examined for potential confounds (e.g. *stadium* might be pronounced with three syllables, or *sibling* having a potential phonological neighbor in *sizzling*).

Procedure

The procedure of Experiment 2 matched that of Experiment 1, with two counterbalanced blocks (one in silence and one requiring concurrent articulation) that had four practice trials and 40 critical trials (20 phonologically-dissimilar and 20 phonologically-similar), followed by an articulation rate measure for both lists. Again, both serial-order recall and free recall scores were calculated even though participants were explicitly instructed to select the answers in order.

Results

As in Experiment 1, the pronunciation duration of the phonologically-dissimilar and phonologically-similar stimuli were compared. There was no significant difference in the speed at which participants recited the randomized 10-item list of phonologically-dissimilar words (*M*

Table 2. Phonologically-similar and phonologically-dissimilar stimuli in Experiment 2.

Word	Length	Freq_HAL	SUBTLWF	OrthoN	PhonoN	UG_Mean	BG_Mean
scalpel	7	203	3.16	0	0	25632.36	1537.33
sausage	7	1273	7.78	0	0	29634.23	900.5
sergeant	8	2405	62.94	0	0	34994.28	2575.14
sulfur	6	678	1.18	0	0	18088.51	751.8
syringe	7	697	1.94	0	0	27158.61	2531
sternum	7	162	0.8	0	0	29041.49	2582.67
sirloin	7	80	0.61	0	0	29092.37	1917.33
saffron	7	390	0.61	0	0	25379.36	1416.33
sequin	6	43	0.14	0	0	26323.24	2003.8
sorbet	6	32	0.27	0	0	31470.12	1202
$M_{Similar}$	**6.8**	**596.3**	**7.94**			**27681.46**	**1741.79**
p-value	0.749	0.48	0.861			0.545	0.361
$M_{Dissimilar}$	**6.9**	**853.5**	**9.58**			**26531.78**	**1514.05**
nostril	7	376	0.69	0	0	30100.29	1867.17
picnic	6	1374	11.69	0	0	22029.17	1385.6
debris	6	1761	3.12	0	0	27672.58	1809.6
trapeze	7	152	1.35	0	0	31365.93	1313
cauldron	8	1019	0.47	0	0	23755.62	1540
lozenge	7	57	0.16	0	0	28268.77	1825.33
musket	6	191	0.98	0	0	24625.94	860.4
thermos	7	306	1.12	0	0	30646.93	1937.67
burglar	7	659	5.53	0	0	19868.72	1266.33
upstairs	8	2640	70.73	0	0	26983.89	1335.43

Note. All data from the English Lexicon Project (Balota et al., 2007) or the Leipzig Corpora Collection (Goldhahn et al., 2012); Freq_HAL = word frequency reported by the HAL study; SUBTLWF = word frequency from Brysbaert & New (2009); OrthoN = number of orthographic neighbors; PhonoN = number of phonological neighbors; UG_Mean = average unigram frequency for all unigrams within a word; BG_Mean = average bigram frequency for all bigrams within a word.

table cont'd.

Word	BGFreqPOS	TG_Mean	I_Mean_RT	Concreteness
scalpel	1092	147.73	803.21	4.86
sausage	1114	151.87	668.78	4.88
sergeant	2271	299.78	838.46	4.7
sulfur	1090	29.03	657.97	4.43
syringe	1571	794.53	839.42	4.81
sternum	1533	437.99	869.73	4.69
sirloin	1958	86.23	760.39	4.66
saffron	1018	176.66	854.69	4.44
sequin	1778	82.42	861.67	4.24
sorbet	1584	114.02	786.43	4.43
$\boldsymbol{M_{Similar}}$	**1500.9**	**232.02**	**794.08**	**4.61**
p-value	0.893	0.655	0.865	0.882
$\boldsymbol{M_{Dissimilar}}$	**1524**	**321.32**	**784.43**	**4.63**
nostril	2131	260.46	705.1	4.89
picnic	1072	52.29	677.48	4.83
debris	1797	138.49	688.79	4.69
trapeze	1843	137.84	833.23	4.55
cauldron	1620	152.22	794.08	4.61
lozenge	1328	99.26	1166.92	4.59
musket	1175	86.31	852.87	4.67
thermos	1333	1954.7	825.31	4.67
burglar	1449	89.88	713.41	4.44
upstairs	1492	241.81	587.15	4.33

Note. BGFreqPOS = sum of the bigram frequency in the same position; TG_Mean = average trigram frequency for all trigrams within a word; I_Mean_RT = mean response time on a lexical decision task.

= 5.50 *s* *SD* = 1.06 s) compared to the randomized 10-item list of phonologically-similar words (*M* = 5.75 s *SD* = 1.04 s), $t(46) = -1.52$, $p = 0.14$.

A visual representation of the PSE in Experiment 2, both with and without silent articulation for both serial-order and free recall scoring can be found in Figure 2. Again, as there were no effects of the counter-balanced block order in either serial- or free-recall conditions, all analysis presented were collapsed across block order. The serial-order recall two-way ANOVA identified a traditional PSE where recall for phonologically-dissimilar words was significantly better than recall for phonologically-similar words, $F(1,49) = 100.42$, $p < 0.001$, $\eta^2_{partial} = .67$. There was also a significant effect of articulation condition where performance in silence was significantly better than performance under silent articulation, $F(1,49) = 108.37$, $p < 0.001$, $\eta^2_{partial} = .69$. However, there was again no significant interaction between phonological similarity and articulation condition, $F(1,49) = 0.56$, $p = 0.43$, $\eta^2_{partial} = .01$. These results indicated that the PSE remained robust even when orthographic and phonological neighbors were eliminated across both lists, and that silent concurrent articulation did significantly reduce the effect.

Once again, the free-recall two-way ANOVA replicated the findings of the serial-order recall ANOVA. There were significant main effects of phonological similarity (dissimilar words > similar words), $F(1,49) = 58.49$, $p < 0.001$, $\eta^2_{partial} = .54$, and articulation condition (silence > silent articulation), $F(1,49) = 65.43$, $p < 0.001$, $\eta^2_{partial} = .58$, and no significant interaction between the two, $F(1,49) = 2.13$, $p = 0.15$, $\eta^2_{partial} < .04$, indicating that observed effects of phonological similarity were likely not due uniquely to order effects.

Discussion

The results of Experiment 2 indicated that controlling for orthographic and phonological neighborhood did not affect the PSE similarly to the WLE. The PSE remained both significant and strong unlike the observed reversal of the WLE in Experiment 1. While the results of Experiment 2 are in line with traditional accounts of rehearsal, the different interactions the effects have with both orthographic and phonological neighborhoods presents problems for their proposed shared cause. As stated above, the PSE can be explained well even in models of

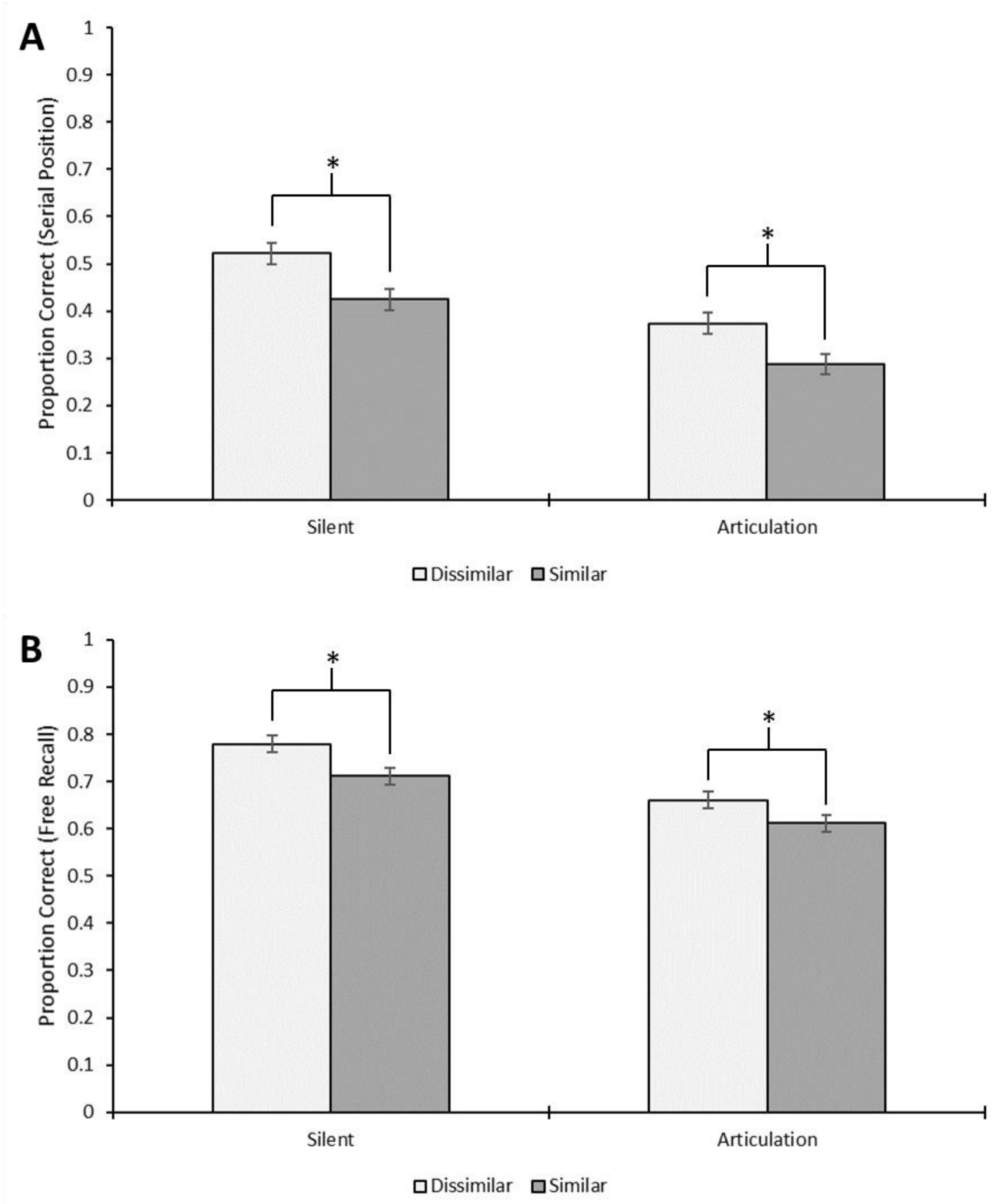

Figure 2. Proportion correct of phonologically-dissimilar and phonologically-similar words under silence and silent concurrent articulation with serial-recall scoring (A) and with free recall scoring (B) in Experiment 2.

memory including no role for rehearsal, but the similar effects of articulatory suppression on the WLE, PSE, and ISE have led to suggesting a shared cause.

It follows that effects sharing a cause should be affected similarly by similar methodological manipulations. However, the present study, in line with McGill and Elliott (in prep), suggests that the WLE and PSE are in-fact not impacted similarly by certain methodologies. McGill and Elliott (in prep), demonstrated that while the PSE was lessened under speeded presentation the WLE was not. These findings suggested that repeated articulation of to-be-remembered words is vital to demonstrating the PSE but not the WLE. Experiment 2 built upon that distinction by demonstrating the reversal of the WLE when both orthographic and phonological neighborhood were controlled upon even though the PSE appeared to be completely unaffected. While it is difficult to say for certain that neighborhood effects cannot impact the PSE without explicit manipulation of neighborhood sizes, it is unlikely that the PSE would be affected as the observed effect in Experiment 2 was particularly strong. While the observed partial eta-squared was lower than some older observations of the PSE in English word stimuli (Baddeley et al., 1984 - $\eta^2_{partial}$ = .91; Coltheart, 1993 - $\eta^2_{partial}$ = .82), more recent examinations are in line with the observed effect size of the PSE whether using English words (Baddeley, Hitch, & Quinlan, 2018 - $\eta^2_{partial}$ = .69) or Dutch words (Lian et al., 2004 - $\eta^2_{partial}$ = .71).

Of additional importance is that the present study demonstrated a significant PSE in free recall and no significant reduction in the size of the PSE under silent concurrent articulation. The persistent effect in free recall may be caused by the decision to manipulate phonological-similarity through only the initial sound of the TBR items. Using TBR items that sound similar across multiple phonemes (e.g. *can*, *cap*, *man*, *map*; Baddeley et al., 1984) may allow for easier recall of the TBR items, because individuals can effectively search memory only for words with those sounds in those positions. However, in Experiment 2, the possible benefit of searching for words beginning with the same initial sound is likely less, as the number of words with any one phoneme as the first phoneme is significantly higher than the number of words that share their first *and* second phonemes. Even when some prior work used stimuli only sharing a single phoneme (e.g. Larsen & Baddeley, 2003; Murray, 1967) the use of letter stimuli suggests that the number of items searched in memory can be significantly limited. Additionally, the use of a reconstruction of order paradigm to recall items, as opposed to a recall procedure that did not supply the items for selection, suggests that individuals do not need to search their entire lexicon to generate a response; instead, they can rely on the 10 presented words as a supplemental cue to recall the words in the correct order. This recall procedure likely lessened participants' need to search their lexicon and generate a response, which would severely lessen any effects of how phonological-similarity was experimentally manipulated in the TBR items. In terms of an explanation for the persistent effect under silent concurrent articulation, while McGill and Elliott (in prep) did demonstrate a significant PSE even under silent concurrent articulation, the effect was significantly less than without articulation. The lack of an observed interaction is somewhat unexpected and might suggest some importance of neighborhood effects to the *elimination* of the PSE under articulation but additional follow-up would be needed.

Experiment 3: The Irrelevant-sound Effect

While the WLE and PSE have been the two most discussed "rehearsal" effects within the literature, the theoretical basis of the ISE also has implications for order processing within rehearsal. However, as discussed earlier, the inability to lessen the size of the ISE through speeded presentation is problematic for the rehearsal interpretation of the effect (McGill & Elliott, in prep). If the ISE is caused by the interference of the order processing of rehearsal by obligatory order processing of auditory stimuli, then the limited ability to sub-vocally recite more than the presently presented item should at the very least lessen the size of the effect. Therefore, it is difficult to assume that order processing in rehearsal is the cause of the ISE. One possible explanation could be that the ISE is caused by the irrelevant sounds interfering with some perceptual order process, but much prior work has demonstrated that the ISE can be demonstrated when irrelevant sounds are presented during a retention interval after the presentation of all the to-be-remembered items (Elliott et al., 2016; Macken et al., 1999; Miles et al., 1991). This pattern of results has been interpreted as strong evidence against the ISE being the result of encoding or perceptual interference alone.

While the ability to temporally separate the stimulus presentation and the irrelevant sounds indicates the ISE is likely not a perceptual effect, the differential effect sizes of irrelevant sounds across a retention interval presents additional issues. Macken et al. (1999) and Elliott et al. (2016) examined the different effects of irrelevant sound presented during different 5000 ms parts of the experimental paradigm. Both identified that the ISE was not significant when presented before the presentation of to-be-remembered items and during the first half of presentation, and a significant ISE was present during the second half of item presentation and both the first and second half of the retention interval (See Figure 3). The lack of an effect at the beginning of item presentation can be accounted for by suggesting that individuals do not need to rehearse cumulatively with only a few items present, but such an explanation could be taken to suggest rehearsal is not preventing the decay of items when only a few are needed to be maintained.

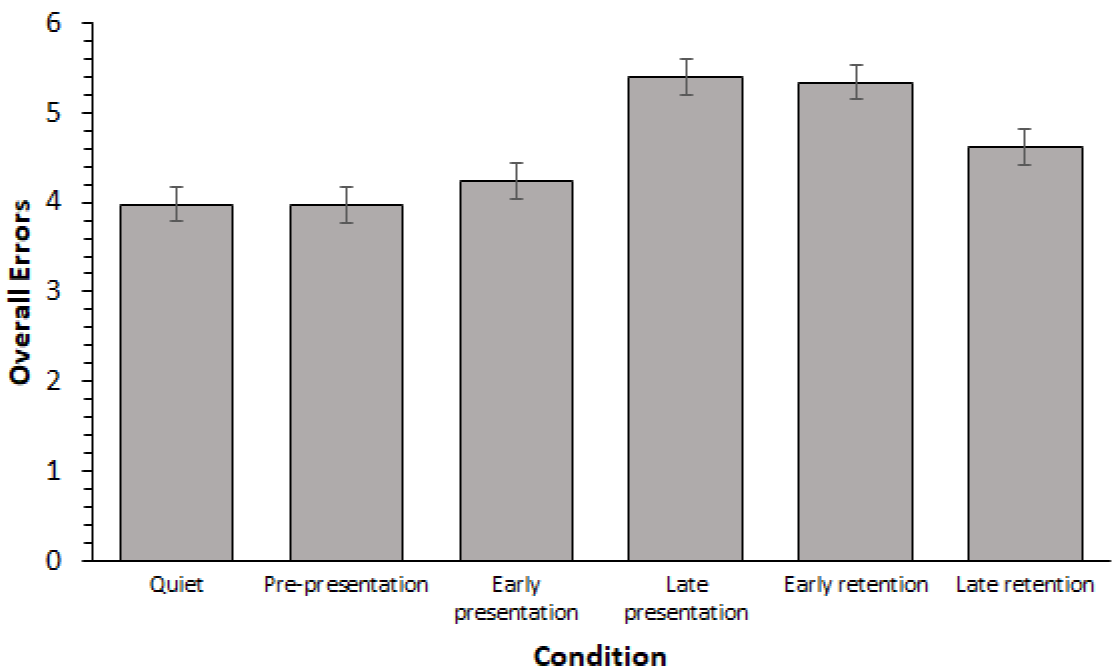

Figure 3. Overall errors in serial order recall for irrelevant sound conditions, adapted from Macken et al. (1999).

However, if the conventional conception of rehearsal is correct, even a small number of items should still be rehearsed in order to keep items in working memory until the recall period. Furthermore, if this is true, even a small number of items should be susceptible to the interference of irrelevant sounds. Critically, the second half of the retention interval in both Macken et al. (1999) and Elliott et al. (2016) had a lessened ISE. While it is suggested that the by the second half of the retention interval the "rehearsal cohort" is stable and no longer susceptible to forgetting, such a suggestion necessarily requires that rehearsal is not used to limit the effects of decay, and instead is used to produce a stable representation of items that does not decay. In memory models including unique short- and long-term memory stores, it would be inefficient and unlikely for individuals to store to-be-remembered items in long-term memory due to an increased amount of inter-trial interference. However, if the items remain in a short-term store, the rate of forgetting due to decay should be the same even during the later portion of the retention interval. Thus, individuals should be equally reliant on rehearsal during the second half of the retention interval as they are in the first half of the retention interval and, therefore, show an equally large effect. As this is not the case, if rehearsal is the cause of the effect, then the role of rehearsal within working memory needs to be adapted to account for rehearsal's importance lessening over time.

One possible explanation is that neighborhood effects impact the ISE in a similar manner to the WLE, and that the effect is lessened over time. While any possible role for neighborhood effects within the ISE is purely speculative, Experiment 3 examines such a possibility. As Experiment 1 clearly demonstrates, there may be a need to parse out effects of rehearsal from effects stemming from either orthographic or phonological neighborhood effects. Even though the reversal of the WLE through controlling neighborhood effects was not replicated in the PSE, there remains a need to determine how and why effects typically attributed to rehearsal can demonstrate differential relationships with methodological manipulations. As prior work with speeded presentation had indicated that the WLE and ISE departed from the PSE as neither were affected by the rehearsal blocking methodology of speeded item presentation (McGill & Elliott, in prep) it could expected that the ISE would demonstrate significant interactions with orthographic and phonological neighborhood similar to the WLE. Additionally, as the ISE is not experimentally manipulated through the to-be-remembered stimuli, Experiment 3 allowed for a fully-crossed examination of neighborhood effects within the ISE where an additional independent variable examined to-be-remembered words with more two or more phonological and orthographic neighbors.

Methods

Participants

Fifty-six Louisiana State University undergraduates aged 17-23 (M_{age} = 18.62, SD_{age} = 1.05) participated for course credit. Two participants did not report their age. Of the 56 participants 48 were female and 8 were male. All participants reported English as their native language, having either normal or corrected vision, and not suffering from any loss of hearing.

Design and power analysis

Experiment 3 employed a within-subjects design with three independent variables. There were three irrelevant sound conditions (silence, steady-state sounds, and changing-state sounds), two articulation conditions (silence and silent concurrent articulation), and two neighborhood size conditions (small and large neighborhood size). Similar to Experiments 1 and 2, the dependent variables were the proportion correct scores using both strict serial-position scoring and free-recall.

Power analysis using BUCSS (Anderson et al., 2017) to control for publication bias and uncertainty suggested a sample size of 42 was needed to replicate the interaction between the ISE and concurrent articulation from Hanley (1997). However, the sample size of 50 was chosen to ensure methodological consistency across Experiments 1-3.

Materials

Neighborhood size was manipulated using the same two-syllable phonologically-dissimilar stimuli from Experiments 1 and 2 for the small neighborhood condition. The large neighborhood condition was created by selecting 10 additional stimuli, all with at least two phonological neighbors and two orthographic neighbors. The additional stimuli were again controlled for word frequency, bigram average, bigram frequency by position, picture naming response time, and concreteness (See *Table 3*). The neighbor condition in Experiment 3 was designed to maximize the number and frequency of the orthographic and phonological neighbors for those items.

Auditory condition was manipulated by using tone auditory stimuli from Elliott (2002) in order to ensure that the auditory stimuli contain no additional neighborhood information that could also interfere with the size of the effect. In the irrelevant sound conditions, a single 250ms tone was presented simultaneously with the onset of each item presentation. The three sound conditions were silence (no auditory stimuli are presented), steady-state (a single repeated tone presented throughout the trial), and changing-state (a different irrelevant sound presented with each to-be-remembered item).

Procedure

The procedure of Experiment 3 was similar to that of Experiments 1 and 2. Experiment 3 again included two counterbalanced blocks of each articulation condition. However, unique to Experiment 3, the blocks consisted of 60 critical trials as opposed to 40. Within each block, all six possible combinations of neighborhood size and irrelevant-sound conditions were randomly presented, with the condition that no trial type was repeated more than once. Additionally, silent

Table 3. Large neighborhood and small neighborhood stimuli in Experiment 3.

Word	Length	Freq_HAL	SUBTLWF	OrthoN	PhonoN	OrthoFreq	PhonoFreq
lotion	6	1147	3.25	3	3	8.97	8.97
sifter	6	16	0.1	4	2	7.17	8.55
mustard	7	1664	6.45	2	3	4.71	5.62
radish	6	139	0.61	2	2	3.65	5.07
kitten	6	2238	4.73	2	6	6.06	8.13
noodle	6	523	2.9	2	8	6.26	6.14
gasket	6	889	0.67	2	2	6.95	6.95
outpost	7	2429	1.31	2	2	3.37	3.37
paddock	7	320	0.33	2	3	5.62	6.5
doorman	7	161	3.18	2	3	4.72	7.75
M_{Large}	**6.4**	**952.6**	**2.35**				
p-value	0.096	0.801	0.31				
M_{Small}	**6.9**	**853.5**	**9.58**				
nostril	7	376	0.69	0	0	0	0
picnic	6	1374	11.69	0	0	0	0
debris	6	1761	3.12	0	0	0	0
trapeze	7	152	1.35	0	0	0	0
cauldron	8	1019	0.47	0	0	0	0
lozenge	7	57	0.16	0	0	0	0
musket	6	191	0.98	0	0	0	0
thermos	7	306	1.12	0	0	0	0
burglar	7	659	5.53	0	0	0	0
upstairs	8	2640	70.73	0	0	0	0

Note. All data from the English Lexicon Project (Balota et al., 2007) or the Leipzig Corpora Collection (Goldhahn et al., 2012); Freq_HAL = word frequency reported by the HAL study; SUBTLWF = word frequency from Brysbaert & New (2009); OrthoN = number of orthographic neighbers; PhonoN = number of phonological neighbors; OrthoFreq = the average word frequency reported by the HAL study of orthographic neighbors; PhonoFreq = the average word frequency reported by the HAL study of phonological neighbors.

table cont'd.

Word	UG_Mean	BG_Mean	BGFreqPOS	TG_Mean	I_Mean_RT	Concreteness
lotion	31039.12	2853	1463	823.87	633.34	4.79
sifter	31786.93	2583.6	2411	412.37	900.77	4.64
mustard	24763.7	1809.83	1916	396.33	654.72	4.93
radish	27531.21	2018.4	1275	239.02	781.03	4.87
kitten	33009.19	2297.2	1875	211.99	611.26	4.86
noodle	30823.67	1197.6	1676	80.82	697.21	4.71
gasket	28584.62	908.8	1324	94.58	683.65	4.44
outpost	27673.69	1355.67	1237	260.28	689.25	4.04
paddock	18621.68	755.67	1077	67.19	692.39	4.22
doorman	27112.48	1669	1575	246.43	685.13	4.79
M_{Large}	**28094.63**	**1744.88**	**1582.9**	**283.29**	**702.88**	**4.63**
p-value	0.395	0.363	0.723	0.848	0.166	0.985
M_{Small}	**26531.78**	**1514.05**	**1524**	**321.32**	**784.43**	**4.63**
nostril	30100.29	1867.17	2131	260.46	705.1	4.89
picnic	22029.17	1385.6	1072	52.29	677.48	4.83
debris	27672.58	1809.6	1797	138.49	688.79	4.69
trapeze	31365.93	1313	1843	137.84	833.23	4.55
cauldron	23755.62	1540	1620	152.22	794.08	4.61
lozenge	28268.77	1825.33	1328	99.26	1166.92	4.59
musket	24625.94	860.4	1175	86.31	852.87	4.67
thermos	30646.93	1937.67	1333	1954.7	825.31	4.67
burglar	19868.72	1266.33	1449	89.88	713.41	4.44
upstairs	26983.89	1335.43	1492	241.81	587.15	4.33

Note. UG_Mean = average unigram frequency for all unigrams within a word; BG_Mean = average bigram frequency for all bigrams within a word; BGFreqPOS = sum of the bigram frequency in the same position; TG_Mean = average trigram frequency for all trigrams within a word; I_Mean_RT = mean response time on a lexical decision task.

concurrent articulation trials were completed in the same manner as in Experiments 1 and 2. After the critical trials, articulation rate measures were assessed for both word lists.

Results

Once again, pronunciation duration for the two groups of stimuli was compared. Contrary to a priori expectations, participants recited the randomized 10-item list of words with orthographic and phonological neighbors (*M* = 6.24 s *std.* = 2.06 s) faster than the randomized

10-item list of words with no orthographic or phonological neighbors (*M* = 5.43 s *SD* = 1.64 s), $t(44) = 3.94$, $p < 0.001$.

A visual representation of the ISE in Experiment 3, both with and without silent articulation for serial-order and free recall scoring can be found in Figure 4 (words with no orthographic or phonological neighbors) and Figure 5 (stimuli with orthographic and phonological neighbors). As there were no effects of the counter-balanced block order in either serial- or free-recall conditions, all analyses presented were collapsed across block order. Experiment 3 employed three-way ANOVAs as opposed to the two-way ANOVAs in Experiments 1 and 2 in order to account for the additional neighborhood size manipulation. The serial-order recall ANOVA identified all three main effects as significant: the main effect of irrelevant sound, $F(2,102) = 8.55$, $p < 0.001$, $\eta^2_{partial} = .14$, where Bonferroni corrected post-hoc comparisons indicated that performance during changing-state irrelevant sound was significantly worse than both performance in silence ($p < 0.001$) and performance during steady-state sounds ($p < 0.05$), and no significant difference between silence and steady-state performance ($p = 0.53$). The main effect of neighborhood, $F(1,51) = 79.40$, $p < 0.001$, $\eta^2_{partial} = .61$, indicated that recall for words with orthographic and phonological neighbors was better than recall for words with no neighbors. The main effect of articulation, $F(1,51) = 78.16$, $p < 0.001$, $\eta^2_{partial} = .61$, indicated that recall was better with no articulation than under silent articulation. All three two-way interactions were significant, articulation and neighborhood size, $F(1,51) = 6.42$, $p < 0.05$, $\eta^2_{partial} = .11$, articulation and irrelevant sound, $F(2,102) = 3.53$, $p < 0.05$, $\eta^2_{partial} = .07$, and neighborhood size and irrelevant sound, $F(1,51) = 7.74$, $p < 0.001$, $\eta^2_{partial} = .13$. However, the three-way interaction did not reach significance, $F(2,102) = 2.97$, $p = 0.06$, $\eta^2_{partial} = .06$.

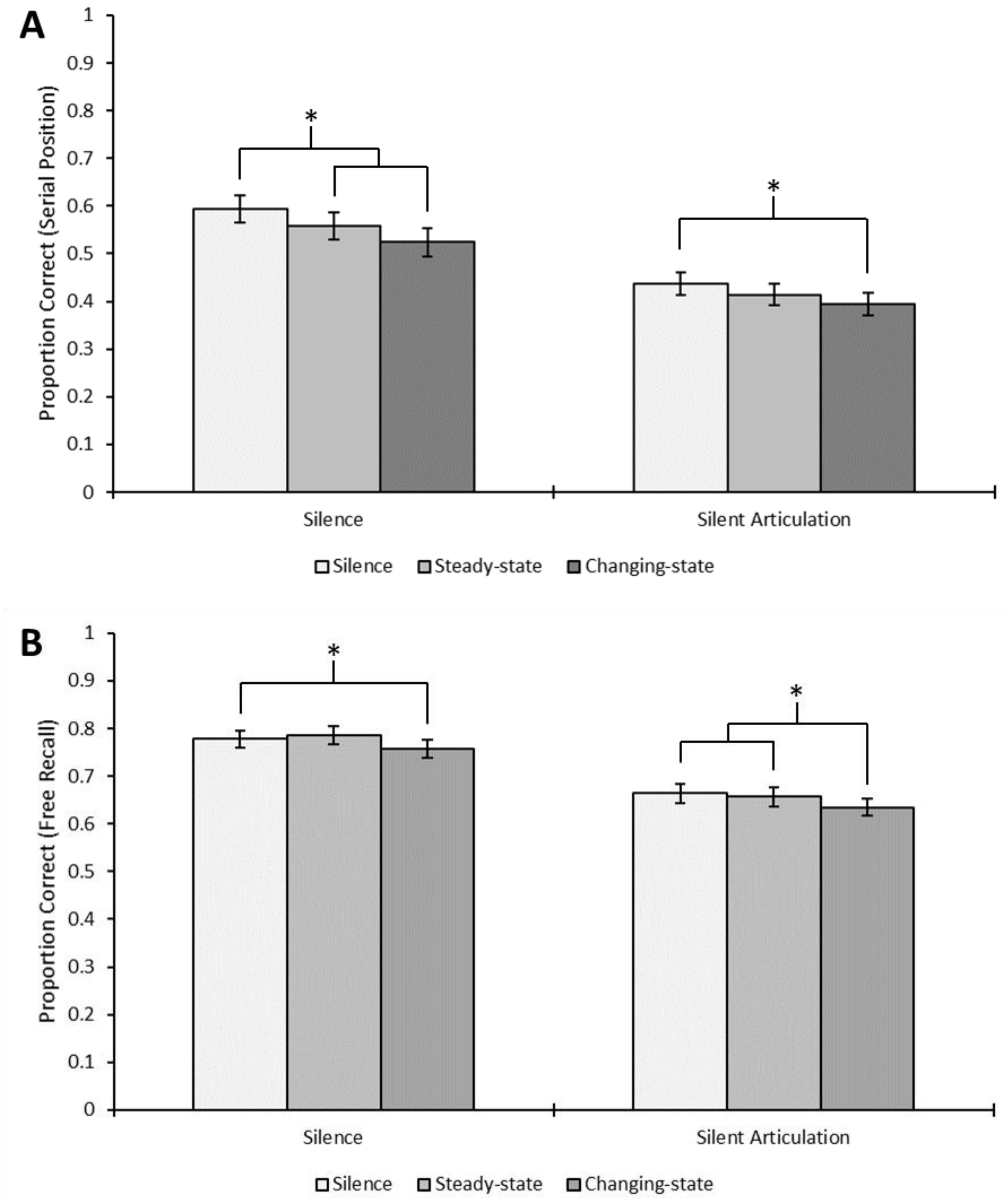

Figure 4. Proportion correct of words with no orthographic or phonological neighbors for different irrelevant sound conditions under silence and silent concurrent articulation with serial-recall scoring (A), and with free recall scoring (B) in Experiment 3.

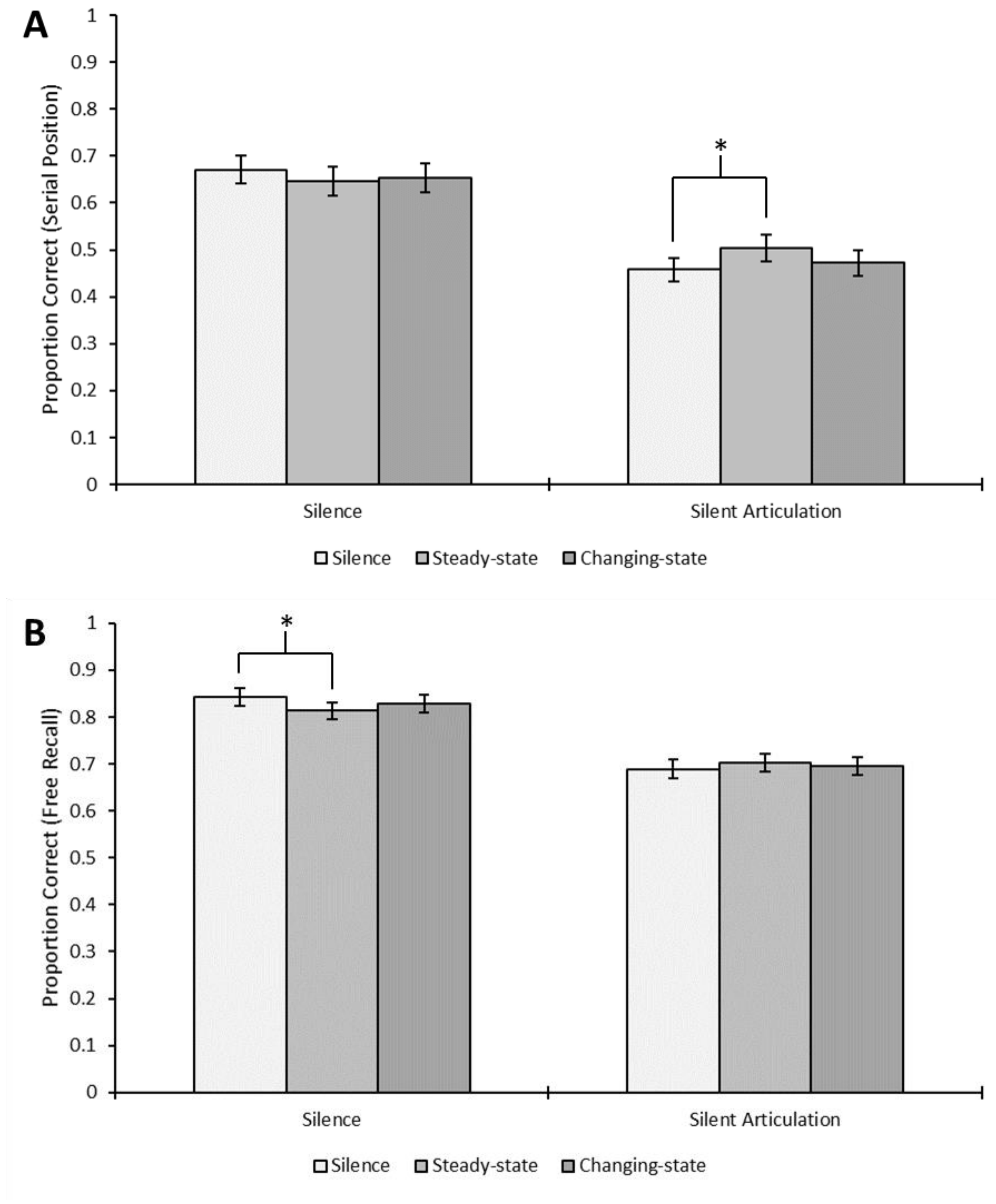

Figure 5. Proportion correct of words with both orthographic and phonological neighbors for different irrelevant sound conditions under silence and silent concurrent articulation with serial-recall scoring (A), and with free recall scoring (B) in Experiment 3.

In order to further investigate the interactions identified, four one-way ANOVAs were performed on the irrelevant sound conditions for each of the neighborhood size by articulation combinations. The first follow-up ANOVA, which was performed on irrelevant sound with no concurrent articulation and no orthographic or phonological neighbors, found a significant main effect, $F(2,102) = 11.50$, $p < 0.001$, $\eta^2_{partial} = .19$. Bonferroni corrected post-hoc analyses identified that recall during both changing-state ($p < 0.001$) and steady-state ($p < 0.05$) sounds was significantly worse than silence, but there was no significant difference between changing- and steady-state sounds ($p = .07$). The next ANOVA examined the effect of irrelevant sounds under silent concurrent articulation for words with no neighbors, and again found a significant main effect of irrelevant sound, $F(2,102) = 6.20$, $p < 0.01$, $\eta^2_{partial} = .11$. However, only recall in silence was significantly greater than recall during changing-state sounds ($p < 0.01$), while steady-state sounds were not significantly different from either changing-state sounds ($p = 0.32$) or silence ($p = 0.19$).

The next set of ANOVAs examined the effects of irrelevant sounds on words with both orthographic and phonological neighbors. With no articulation, the present observation found no effects of irrelevant sounds on performance, $F(2,102) = 1.33$, $p = 0.27$, $\eta^2_{partial} = .03$. indicating that without silent articulation, words with both orthographic and phonological neighbors produced no significant effects of irrelevant sound on serial order recall. However, when recall for words with neighbors under concurrent silent articulation was examined, there was a significant main effect, $F(2,102) = 4.88$, $p < 0.01$, $\eta^2_{partial} = .09$. Uniquely, the significant effect was driven by recall performance in silence being significantly worse than performance during steady state sounds ($p < 0.01$). There were no significant differences between changing-state sounds and either silence ($p = 1.00$) or steady-state sounds ($p = 0.11$).

The second three-way ANOVA examined performance using free recall scoring for irrelevant sound, neighborhood, and articulation conditions. Significant main effects on free recall were identified for articulation (no articulation > silent articulation), $F(1,51) = 72.68$, $p < 0.001$, $\eta^2_{partial} = .59$, neighborhood (words with neighbors > words with no neighbors), $F(1,51) = 52.23$, $p < 0.001$, $\eta^2_{partial} = .51$, and irrelevant sound, $F(2,102) = 3.77$, $p < 0.05$, $\eta^2_{partial} = .07$. Post-hoc comparisons indicated that the irrelevant sound main effect was driven by free recall in silence being significantly greater than in changing-state sounds ($p < 0.05$), as there was no significant difference between free recall during steady-state sounds and either changing-state sounds ($p = 0.15$) or silence ($p = 1.00$). Unlike the serial order recall ANOVA neither the articulation by neighborhood, $F(1,51) = 1.27$, $p = 0.26$, $\eta^2_{partial} = .02$, nor the articulation by sound, $F(2,102) = 0.88$, $p = .42$, $\eta^2_{partial} = .02$, interactions were significant. There was a significant neighborhood by irrelevant sound interaction, $F(2,102) = 4.73$, $p < 0.05$, $\eta^2_{partial} = .09$, and unlike in serial recall scoring, a significant three-way interaction, $F(2,102) = 3.47$, $p < 0.05$, $\eta^2_{partial} = .06$, was found.

Similar follow-up ANOVAs on irrelevant sound conditions for each combination of articulation and neighborhood were performed on the free recall scores. For the no articulation and no orthographic or phonological neighbor conditions a significant main effect of irrelevant sound was identified, $F(2,102) = 3.36$, $p < 0.05$, $\eta^2_{partial} = .06$, which was driven by a free recall being worse during changing-state sounds than during steady-state sounds ($p < 0.05$). Neither changing-state ($p = 0.21$) nor steady-state ($p = 1.00$) sounds significantly altered free recall performance when compared to silence. The silent articulation condition appeared to have a minimal effect on free recall for words with no orthographic neighbors, $F(2,102) = 5.60$, $p < 0.01$, $\eta^2_{partial} = .10$. However, free recall performance was significantly worse for changing-state sounds than for both silence ($p < 0.01$) and steady-state sounds ($p < 0.05$), while there was still no significant difference between silence and steady-state sounds ($p = 1.00$).

Examining the effects of irrelevant sound on the free recall of words with orthographic and phonological neighbors without an articulation requirement found another significant main effect, $F(2,102) = 3.83$, $p < 0.05$, $\eta^2_{partial} = .07$. Performance under steady-state sounds was significantly worse than in silence ($p < 0.05$), while there was no significant differences between changing-state sounds and either steady-state sounds ($p = .41$) or silence ($p = 0.64$). In the articulation condition for words with orthographic and phonological neighbors, there was no identified effects of irrelevant sounds on free recall, $F(2,102) = 0.73$, $p = 0.48$, $\eta^2_{partial} = .01$.

Discussion

Experiment 3 demonstrated just how important the investigation of simultaneous neighborhood effects during “rehearsal” effects might be. To start, there was no evidence to support the traditional effects of irrelevant sound on serial order recall in words that were determined to have both orthographic and phonological neighbors. While there was a significant main-effect in the silent articulation condition, it was caused by significantly worse recall in the silent condition than in the steady-state irrelevant sound condition. Not only was there not a traditional ISE, the steady-state effect was reversed and the changing-state effect was reversed in magnitude even if it was statistically non-significant. While it is possible that the reversed steady-state effect is a statistical anomaly, the lack of a traditional effect in either articulation condition presents striking problems for rehearsal accounts of the ISE.

It is not presently clear what mechanisms may be driving the elimination of the ISE in the neighborhood condition, because there is no expectation of any such interaction under an ISE model, suggesting the cause is order interference during sub-vocal item recitation. It would be impossible to suggest that no prior work in the ISE used TBR stimuli with neighbors, but it may be that the specific combination of methodological choices in the Experiment 3 presented some unique opportunity. Different factors to consider may be the use of words as the TBR stimuli, the use of tones as the to-be-ignored stimuli (which sometimes lead to smaller effect sizes than the use of speech; e.g. Elliott, 2002), the use of a closed set of unrelated TBR stimuli, the reconstruction recall requirements, and/or other methodological factors. For example, contrary to the methodology of the current Experiment 3, it is common to examine the ISE using individual letter (e.g. Bell, Röer, & Buchner, 2013; Larsen & Baddeley, 2003) or number (e.g. Elliott, 2002; Hughes et al., 2007) stimuli in order to maximize capacity and lessen the maintenance requirement of item information for participants.

As an additional methodological consideration, the use of simplistic TBR stimuli in an ISE paradigm allows experimenters to use words as the irrelevant sound conditions while not having to control or account for possible semantic interference between TBR and to-be-forgotten items (e.g., Marsh, Hughes, & Jones, 2009). It is possible that certain stimulus sets may limit the efficacy of neighborhood activation for later recall, and therefore, individuals would no longer rely on that activation for maintenance. For example, it is possible that when using closed sets of related stimuli(e.g. digits, letters, days of the week), individuals can limit their search to only the items in the set, and thus do not rely on the neighborhood activation to reconstruct a memory trace. Furthermore, the recall method used presently, in which participants selected six out of ten possible stimuli, might allow individuals to rely on unique strategies that would otherwise be less effective. The specific combination of methodological choices that were made prior to running a single participant in Experiment 3 may have provided a unique opportunity for the ISE to be eliminated by words with orthographic and phonological neighbors. Further systematic variation of these methodological choices may provide additional information to evaluate such an assumption. For example, Experiment 3 could easily be replicated requiring un-cued serial-order recall, with TBR stimuli that all share semantic relationships, or even with irrelevant sound conditions comprising English words where neighborhood size of the irrelevant items is manipulated. Any of the three manipulations described above may provide useful insight into the cause of the unique and unexpected findings of the ISE being eliminated when TBR items have both orthographic and phonological neighbors. However, until proper investigation occurs it would be reckless to speculate on the exact reason.

Furthermore, just as in the WLE, it is possible that some of the differential and/or conflicting results identified in prior ISE work might be attributable to stimulus selection for either the TBR items, or possibly the irrelevant sound stimuli. As discussed above, it is not uncommon for irrelevant sound conditions to comprise word stimuli, as opposed to the tones used in Experiment 3. If neighborhood size does impact the ISE, it stands to reason that the neighborhood size of the TBR stimuli might also interact with the effect size. Future examinations may even identify unique effects of neighborhood on TBR and irrelevant stimuli, and leverage that information into a better understanding of how and why the ISE is impacted by orthographic and phonological neighborhood. If successful, understanding these neighborhood effects may be the key to understanding why the steady-state effect might occur in some situations and not others.

Returning to the measurement of articulation time for the stimuli, there may be an inclination to attribute the results of Experiment 3 to the increased articulation time for the TBR items without neighbors. However, making such a conclusion would conflict with the results of Experiment 1. As longer articulation duration would, if anything, improve the recall of items when neighborhood size was controlled, the lack of an ISE in the shorter duration items that also happened to have neighbors further implicates neighborhood size as being the true cause of the findings in Experiment 3. Additional support comes from the known increase in lexical access associated with increased neighborhood size (Andrews, 1989; Sears et al., 1995). This increased lexical access may result in participants being able to more quickly read and recite stimuli with neighbors more quickly.

The elimination of the traditional ISE when using words with orthographic and phonological neighbors was not the only peculiarity observed in Experiment 3. Possibly the most unexpected finding was the ability to observe effects of irrelevant sounds both during silent concurrent articulation and when using free-recall scoring. The ability to eliminate the ISE with concurrent articulation is well established (Hanley, 1997; McGill & Elliott, in prep), and McGill and Elliott (in prep) even demonstrated the ability to eliminate the effects of irrelevant sound using the silent concurrent articulation paradigm used presently. It is unclear why the effects of irrelevant sound were not eliminated in Experiment 3, but it is important to note the reduction in the effect size caused by silent articulation in the no neighborhood condition. Again, it is possible the stimulus selection of TBR items and/or irrelevant sounds was responsible for these differences, but without future examination it is difficult to make any meaningful conclusions about *why* the results differ.

Similarly, the effects of irrelevant sound persisted in most instances of free recall scoring. While the overall effect sizes were relatively small when compared to serial order recall, the ability to demonstrate non-order-based differences in recall in the ISE was not an anticipated finding. Traditionally, if order processing is removed or deemphasized from the recall task, there are no effects of irrelevant sounds (Beaman & Jones, 1997). However, prior work has suggested that when participants use an order-based strategy to maintain items, irrelevant sounds can cause the loss of item information (Beaman & Jones, 1998) which would result in worsened free recall. So, while the mechanism causing the ISE may not affect item memory directly, by reducing the efficacy of order cues used at recall, individuals are also slightly less likely to identify the items. Again, it is entirely possible that neighborhood effects may contribute to these findings and/or assumptions; however, speculation on how or why would be negligent without proper follow-up examination.

General Discussion

The results of all three experiments reinforced the notion that standard models of rehearsal are unable to account for the WLE, PSE, and ISE in its present form. Both the WLE and ISE demonstrated significant effects of orthographic and phonological neighborhood that are difficult to impossible to account for as an effect of traditional rehearsal. First, the WLE was completely reversed by controlling across long and short stimuli for neighborhood effects, directly refuting Guitard et al. (2017). While this direct conflict in the findings is somewhat problematic, three factors play into the present conclusion that Experiment 1 represented a reliable finding. First, Experiment 1 controlled for the same variables as Guitard et al. (2017), with the additional explicit controls like concreteness and phonological neighborhood size. Second, Experiment 1 directly replicated the results of Jalbert, Neath, and Surprenant (2011), and indirectly replicated the findings of both Jalbert, Neath, Bireta, and Surprenant (2011) and Derraugh, Neath, Surprenant, Beaudry, & Saint-Aubin (2017). Third, extensive prior literature demonstrated the importance of stimulus selection in explaining initially conflicting results within the PSE literature (e.g., Bireta et al., 2006). While further examination is clearly needed[1], the above three points can be taken together as strong support for the WLE being truly confounded by effects of orthographic and phonological neighborhood.

---

[1] Guitard, Gabel, Saint-Aubin, Surprenant, and Neath (2018) recently demonstrated the important effects that lexical controls can have on the WLE. Through multiple experiments, it was demonstrated that increasing the number of lexical controls within the WLE resulted in fewer significant differences. Specifically, when concreteness, imageability, familiarity, word frequency, orthographic neighborhood size, frequency of orthographic neighbors, and contextual diversity were controlled for, the WLE was eliminated, but differences in neighborhood size were not.

The results of Experiment 1 and Experiment 3 differed from Experiment 2. While not as clear as in Experiment 1, the effects of neighborhood size in the ISE in Experiment 3 also presented significant problems for traditional rehearsal explanations of the effect. The pattern of findings suggested that TBR items are less susceptible to the order interference of irrelevant sounds when they have orthographic and phonological neighbors. No model of the ISE currently can account for such a possibility without significant revision. There would be no reason to expect the decay of order information that is hypothesized to be prevented in a traditional rehearsal model (Baddeley, 1986; 2000; Camos et al., 2009; Cowan, 2005) would be affected by neighborhood size. While it is not unreasonable to hypothesize that the additional relationships established when TBR items have neighbors makes it easier for individuals to reconstruct a memory trace accurately, how and why those effects would or would not be impacted by irrelevant sound remains a significant question. Furthermore, this marks at least the second time that the WLE and ISE have been shown to depart from the expectations of a traditional rehearsal model in a similar way. The lack of an interaction between either the WLE or ISE and speeded presentation (McGill & Elliott, in prep.) also presents problems.

Returning to Experiment 2, in contrast to the WLE and ISE, the PSE demonstrated a standard effect of similarity and a typical effect size when neighborhood effects were controlled. While the persistence of the PSE alone would, in a vacuum, suggest that rehearsal may be the cause, the deviation from the patterns of results of the WLE and ISE again presents problems for traditional rehearsal models. Like above, this is another example of the PSE conforming to the expectations of a traditional rehearsal model, while the WLE and ISE do not (McGill & Elliott, in prep.). This distinction of the PSE from the other two effects could be considered problematic for traditional models of rehearsal as well. While the PSE can be well explained both with and without the need for sub-vocal rehearsal, the similar effects of concurrent articulation on it, the WLE, and the ISE was believed to support all three sharing a common cause. At the time, the seemingly best explanation was the sub-vocal recitation of TBR items in order to prevent decay. However, if the PSE does significantly depart from the WLE and ISE in cause, there is no need to suggest sub-vocal rehearsal as the cause over alternative explanations (e.g. increased inter-item interference; Nairne 1990). That is not to say that the PSE is not an effect caused by rehearsal, but that other possibilities become more likely.

Furthermore, these results explicitly highlight the need to replicate any potential findings before any final conclusions can be drawn. While prior work in the WLE can be held up as a specific example for the importance of replicating results with a wide number of stimulus sets, the PSE and ISE must be further examined. In the PSE, a single lack of an interaction with neighborhood effects in an uncrossed experimental design should not be considered definitive proof. In the ISE, the apparent finding that using TBR items with neighbors eliminates the effect could only be true under one of two conditions: 1) all prior work finding significant effects using stimuli with neighbors was incorrect, or 2) some additional unexamined variable plays a role in when and how the ISE interacts with neighborhood effects. While the former is highly unlikely, both possibilities can be examined through replication of Experiment 3. In fact, in the present study, there is some suggestion that there may be additional important variables unaccounted for in the design.

Specifically, to determine the role that any individual stimulus played in performance across the experiments, analyses were done on each word (See Appendices A-C for full comparisons of proportion correct in serial-order and free-recall per word). While there was some natural variation in the likelihood an individual correctly recalled any particular word across all three experiments, the word *trapeze* was less likely to be recalled in correct serial position than other 2-syllable, dissimilar, no neighbor words across all three experiments. In fact, across all six stimulus sets, the serial position proportion correct scores were 0.128(Exp. 1), 0.101(Exp. 3), and 0.059 (Exp. 2) lower than the average proportion correct of all similar items. Even though no inferential statistics were performed, only one other word, *elevator* (0.059), showed a similar drop in proportion correct in serial order scoring. While *trapeze* does not stand out from any other words in the explicit controls of Experiments 1-3, it is possible that *trapeze* is indicative of another uncontrolled variable important to recall and/or rehearsal.

Overall, while the present experiments present significant issues for the decay model of rehearsal, these findings disprove neither decay nor rehearsal individually. For decay specifically, while many models of memory including a role for decay are based on the assumption that rehearsal specifically combats the effects of decay, it is possible that novel experimentation will provide a better understanding of if/how time affects memory processes. In regards to rehearsal, despite the tendency of interference only models of memory to suggest rehearsal as a proxy-effect, there remains a strong possibility that rehearsal plays an active role in memory even if not as a decay-prevention mechanism. As most individuals seem to automatically rehearse TBR items without prompting, and do so (at least meta-cognitively) to improve their later recall, to immediately dismiss rehearsal as a mechanism because the decay model is not supported might be rash. However, while neither decay nor rehearsal alone can be dismissed, the present models tying the two together seem unable to adequately account for the findings of the three present experiments without significant revision.

While it is difficult to speculate exactly what is to come from further examination of neighborhood effects and rehearsal, one possibility is that the two are not as unique as they initially appeared. It is possible to propose a model of sub-vocal rehearsal that explains all of these findings at one time: (1) the WLE, PSE, and ISE while accounting for the present results, (2) McGill and Elliott (in prep.), and (3) most contemporary literature in each of the effects. It can be hypothesized that the sub-vocal recitation of information is not done to prevent the decay of TBR items, as often suggested (e.g., Baddeley, 2002; Cowan, 2005), but instead to activate the orthographic and phonological neighbors of the TBR word. Such a mechanism could explain why the WLE is confounded by neighborhood effects as longer words in general have fewer neighbors to activate and are normally recalled more poorly; however, when long and short words have no neighbors, longer words benefit from having more cues to reconstruct the memory trace. In this case, when an individual sub-vocally recites the TBR stimuli and that item has neighbors to activate, the rehearsal process facilitates that activation. Thus, the ability to sub-vocalize is vital to the effect commonly identified as the WLE, even if the beneficial recall for short words is actually caused by neighborhood effects.

In the PSE, inter-item interference causes TBR items to be recalled worse when they are phonologically similar, and phonologically similar words may be more likely to share neighbors or be neighbors with one another, introducing additional interference. Thus, phonological-similarity has an inherent negative effect on the ability to recall items due to there being fewer distinct characteristics associated with each. However, when words share phonology the rehearsal process may be activating similar orthographic and phonological neighbors across TBR items, which would result in additional inter-item interference and lessen recall even more. In Experiment 2, since no words ever had any neighbors, no additional interference would ever be expected, but future examination may be able to identify such an interaction. In the ISE, the recitation of TBR items activates neighbors, which then are used to help establish additional order information, and the result is that individuals become more resistant to the order interference of the irrelevant sounds. When stimuli have no neighbors, the order interference of the irrelevant sounds significantly harms recall. In Experiment 3, the TBR items with orthographic and phonological stimuli may not have shown any significant effects because rehearsal activated the neighbors of the TBR items, repeatedly, in order. Thus, during recall participants were able to reconstruct the order of the TBR items with not only the normal order cues established between the items themselves, but also the order cues established between one item's activated neighbors and another item's activated neighbors *or* one item's activated neighbors and another item itself.

Such a model may also explain why speeded presentation only affected the PSE. In the WLE, a single sub-vocal recitation could be expected to activate some neighbors and benefit the shorter words disproportionately. In the PSE, the speeded presentation does not allow for the sub-vocal recitation to introduce additional cross item interference for words sharing neighbors. In the ISE, speeded presentation should have either no effect on items that demonstrate a significant effect under normal presentation or increase the size of the effect as any combating of the effect from neighbors is likely to be lessened without the ability to repeatedly sub-vocalize and activate those neighbors. However, by that same logic it would be reasonable to assume that speeded presentation should reintroduce a significant effect for items with neighbors.

Taken together, all three experiments present numerous questions. There is no obvious model of sub-vocal rehearsal or any other mechanism that can be applied to explain all three effects, and it is difficult to say for certain whether or not the three effects are related to one another or not. While that may appear to be a negative, it is in fact a positive step towards a better understanding of memory systems. Sub-vocal rehearsal may be one of the most prevalent mechanisms within models of memory, while also often times being an afterthought. However, if the experimental effects that we use to understand and make predictions about rehearsal's role within memory do not stand up to thorough examination, it is vital that we identify how and why those common assumptions are wrong. It may be that only slight modification is needed, clarifying when and how rehearsal can work as a supplement to other mechanisms, or it may be that significant changes to models of memory will be required. Without confidence in the accuracy of even the most basic of assumptions, we cannot be confident in any model of memory.

Appendix A. Proportion Correct Per Word in Experiment 1

Word	Condition	SP Avg	FR Avg	SP Diff	FR Diff
automobile	4-syllable	0.578	0.734	0.013	-0.040
brontosaurus	4-syllable	0.600	0.830	0.035	0.055
elevator	4-syllable	0.507	0.698	-0.059	-0.077
formaldehyde	4-syllable	0.614	0.822	0.049	0.047
geologist	4-syllable	0.562	0.785	-0.003	0.010
kaleidoscope	4-syllable	0.569	0.800	0.004	0.025
meteoroid	4-syllable	0.557	0.749	-0.008	-0.026
terracotta	4-syllable	0.561	0.836	-0.004	0.061
ukulele	4-syllable	0.595	0.791	0.030	0.016
videotape	4-syllable	0.510	0.702	-0.056	-0.073
burglar	2-syllable	0.511	0.697	0.008	-0.039
cauldron	2-syllable	0.544	0.761	0.041	0.025
debris	2-syllable	0.465	0.697	-0.038	-0.038
lozenge	2-syllable	0.560	0.804	0.057	0.069
musket	2-syllable	0.514	0.805	0.011	0.069
nostril	2-syllable	0.481	0.725	-0.022	-0.011
picnic	2-syllable	0.503	0.716	0.000	-0.020
thermos	2-syllable	0.511	0.790	0.008	0.055
trapeze	2-syllable	0.375	0.625	-0.128	-0.111
upstairs	2-syllable	0.566	0.735	0.063	0.000

Note. SP Avg = average serial order recall controlled for serial position; FR Avg = average proportion correct for free recall controlled for serial position; SP Diff = the difference from serial order recall average within the condition; FR Diff = the difference from free recall average within the condition.

Appendix B. Proportion Correct Per Word in Experiment 2

Word	Condition	SP Avg	FR Avg	SP Diff	FR Diff
saffron	Similar	0.374	0.679	0.019	0.020
sausage	Similar	0.345	0.630	-0.009	-0.028
scalpel	Similar	0.356	0.671	0.001	0.013
sequin	Similar	0.415	0.714	0.060	0.056
sergeant	Similar	0.347	0.726	-0.008	0.068
sirloin	Similar	0.357	0.675	0.002	0.016
sorbet	Similar	0.361	0.662	0.007	0.004
sternum	Similar	0.331	0.644	-0.023	-0.014
sulfur	Similar	0.320	0.573	-0.035	-0.085
syringe	Similar	0.340	0.609	-0.015	-0.049
burglar	Dissimilar	0.429	0.652	-0.011	-0.060
cauldron	Dissimilar	0.433	0.699	-0.007	-0.012
debris	Dissimilar	0.457	0.721	0.017	0.009
lozenge	Dissimilar	0.474	0.754	0.033	0.042
musket	Dissimilar	0.413	0.729	-0.027	0.017
nostril	Dissimilar	0.437	0.729	-0.003	0.017
picnic	Dissimilar	0.441	0.711	0.001	-0.001
thermos	Dissimilar	0.443	0.734	0.002	0.023
trapeze	Dissimilar	0.381	0.655	-0.059	-0.057
upstairs	Dissimilar	0.494	0.733	0.054	0.021

Note. SP Avg = average serial order recall controlled for serial position; FR Avg = average proportion correct for free recall controlled for serial position; SP Diff = the difference from serial order recall average within the condition; FR Diff = the difference from free recall average within the condition.

Appendix C. Proportion Correct Per Word in Experiment 3

Word	Condition	SP Avg	FR Avg	SP Diff	FR Diff
doorman	Neighbors	0.542	0.744	-0.027	-0.024
gasket	Neighbors	0.516	0.732	-0.052	-0.037
kitten	Neighbors	0.600	0.772	0.032	0.003
lotion	Neighbors	0.591	0.829	0.022	0.061
mustard	Neighbors	0.547	0.773	-0.021	0.005
noodle	Neighbors	0.560	0.755	-0.009	-0.014
outpost	Neighbors	0.574	0.755	0.005	-0.014
paddock	Neighbors	0.571	0.791	0.003	0.023
radish	Neighbors	0.570	0.757	0.001	-0.011
sifter	Neighbors	0.615	0.775	0.046	0.007
burglar	No neighbors	0.475	0.665	-0.012	-0.056
cauldron	No neighbors	0.518	0.741	0.031	0.020
debris	No neighbors	0.496	0.715	0.010	-0.006
lozenge	No neighbors	0.531	0.779	0.044	0.058
musket	No neighbors	0.473	0.743	-0.013	0.022
nostril	No neighbors	0.485	0.723	-0.002	0.002
picnic	No neighbors	0.497	0.711	0.010	-0.010
thermos	No neighbors	0.451	0.769	-0.035	0.048
trapeze	No neighbors	0.385	0.647	-0.101	-0.074
upstairs	No neighbors	0.554	0.718	0.068	-0.003

Note. SP Avg = average serial order recall controlled for serial position; FR Avg = average proportion correct for free recall controlled for serial position; SP Diff = the difference from serial order recall average within the condition; FR Diff = the difference from free recall average within the condition.

Appendix D. IRB Approval

ACTION ON EXEMPTION APPROVAL REQUEST

LSU

Institutional Review Board
Dr. Dennis Landin, Chair
130 David Boyd Hall
Baton Rouge, LA 70803
P: 225.578.8692
F: 225.578.5983
irb@lsu.edu
lsu.edu/research

TO: Emily Elliott
Psychology

FROM: Dennis Landin
Chair, Institutional Review Board

DATE: May 19, 2017

RE: **IRB#** E2060

TITLE: An Investigation of Cognitive Performance

New Protocol/Modification/Continuation: Modification

Brief Modification Description: Ask participants to provide self-reported GPA, obtain their permission to obtain their actual GPAs, and revise consent reflecting this change.

Review date: 5/18/2017

Approved ___X___ **Disappproved** __________

Approval Date: 5/19/2017 **Approval Expiration Date:** 2/2/2020

Re-review frequency: (three years unless otherwise stated)

LSU Proposal Number (if applicable):

Protocol Matches Scope of Work in Grant proposal: (if applicable)

By: Dennis Landin, Chairman ___DLandin___

PRINCIPAL INVESTIGATOR: PLEASE READ THE FOLLOWING –
Continuing approval is CONDITIONAL on:

1. Adherence to the approved protocol, familiarity with, and adherence to the ethical standards of the Belmont Report, and LSU's Assurance of Compliance with DHHS regulations for the protection of human subjects*
2. Prior approval of a change in protocol, including revision of the consent documents or an increase in the number of subjects over that approved.
3. Obtaining renewed approval (or submittal of a termination report), prior to the approval expiration date, upon request by the IRB office (irrespective of when the project actually begins); notification of project termination.
4. Retention of documentation of informed consent and study records for at least 3 years after the study ends.
5. Continuing attention to the physical and psychological well-being and informed consent of the individual participants including notification of new information that might affect consent.
6. A prompt report to the IRB of any adverse event affecting a participant potentially arising from the study.
7. Notification of the IRB of a serious compliance failure.
8. **SPECIAL NOTE: Make sure you use bcc when emailing more than one recipient. Approvals will automatically be closed by the IRB on the expiration date unless the PI requests a continuation.**

**All investigators and support staff have access to copies of the Belmont Report, LSU's Assurance with DHHS, DHHS (45 CFR 46) and FDA regulations governing use of human subjects, and other relevant documents in print in this office or on our World Wide Web site at http://www.lsu.edu/irb*

References

Anderson, S. F., Kelley, K., & Maxwell, S. E. (2017). Sample-size planning for more accurate statistical power: A method adjusting sample effect sizes for publication bias and uncertainty. *Psychological Science (0956-7976)*, *28*(11), 1547-1562.

Andrews, S. (1989). Frequency and neighborhood effects on lexical access: Activation or search? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(5), 802-814. doi:10.1037/0278-7393.15.5.802

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

Baddeley, A. D., Chincotta, D., Stafford, L., & Turk, D. (2002). Is the word length effect in STM entirely attributable to output delay? Evidence from serial recognition. *Quarterly Journal of Experimental Psychology: Section A*, *55*(2), 353-369.

Baddeley, A. D., Hitch, G. J., & Quinlan, P. T. (2018). Is the phonological similarity effect in working memory due to proactive interference? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(8), 1312–1316. https://doi-org.libezp.lib.lsu.edu/10.1037/xlm0000509

Baddeley, A. D., & Hull, A. (1979). Prefix and suffix effects: Do they have a common basis? *Journal of Verbal Learning & Verbal Behavior*, *18*(2), 129-140. doi:10.1016/S0022-5371(79)90082-3

Baddeley, A. D., Lewis, V., & Vallar, G. (1984). Exploring the articulatory loop. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, *36A*(2), 233-252. doi:10.1080/14640748408402157

Baddeley, A. D., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning & Verbal Behavior*, *14*(6), 575-589. doi:10.1016/S0022-5371(75)80045-4

Beaman, C. P., & Jones, D. M. (1997). Role of serial order in the irrelevant speech effect: Tests of the changing-state hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*(2), 459-471. doi:10.1037/0278-7393.23.2.459

Beaman, C. P., & Jones, D. M. (1998). Irrelevant sound disrupts order information in free recall as in serial recall. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, *51A*(3), 615–636. doi:10.1080/027249898391558

Bell, R., Röer, J. P., & Buchner, A. (2013). Irrelevant speech disrupts item-context binding. Experimental Psychology, 60(5), 376–384. doi:10.1027/1618-3169/a000212

Bireta, T. J., Neath, I., & Surprenant, A. M. (2006). The syllable-based word length effect and stimulus set specificity. *Psychonomic Bulletin & Review*, *13*(3), 434-438. doi:10.3758/BF03193866

Buschke, H. (1963). Relative retention in immediate memory determined by the missing scan method. *Nature, 200,* 1129-1130. doi:10.1038/2001129b0

Camos, V., Lagner, P., & Barrouillet, P. (2009). Two maintenance mechanisms of verbal information in working memory. *Journal of Memory and Language*, *61*(3), 457-469. doi:10.1016/j.jml.2009.06.002.

Caplan, D., Rochon, E., & Waters, G. S. (1992). Articulatory and phonological determinants of word length effects in span tasks. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, *45A*(2), 177-192. doi:10.1080/14640749208401323Colle & Welsh, 1976

Colle, H. A., & Welsh, A. (1976). Acoustic masking in primary memory. *Journal of Verbal Learning & Verbal Behavior*, *15*(1), 17-31. doi:10.1016/S0022-5371(76)90003-7.

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*(4), 539-545. doi:10.3758/BF03197185

Coltheart, V., & Langdon, R. (1998). Recall of short word lists presented visually at fast rates: Effects of phonological similarity and word length. *Memory & Cognition*, *26*(2), 330-342. doi:10.3758/BF03201144

Cowan, N. (2005). *Working memory capacity*. Hove, East Sussex, UK: Psychology Press.

Cowan, N., Baddeley, A. D., Elliott, E. M., & Norris, J. (2003). List composition and the word length effect in immediate recall: A comparison of localist and globalist assumptions. *Psychonomic Bulletin & Review*, *10*(1), 74-79. doi:10.3758/BF03196469

Cowan, N., Nugent, L. D., & Elliott, E. M. (2000). Memory-search and rehearsal processes and the word length effect in immediate recall: A synthesis in reply to Service. *Quarterly Journal of Experimental Psychology: Section A*, *53*(3), 666-670.

Derraugh, L. S., Neath, I., Surprenant, A. M., Beaudry, O., & Saint-Aubin, J. (2017). The effect of lexical factors on recall from working memory: Generalizing the neighborhood size effect. *Canadian Journal of Experimental Psychology*, *71*(1), 23–31. https://doi-org.libezp.lib.lsu.edu/10.1037/cep0000098

Elliott, E. M. (2002). The irrelevant-speech effect and children: Theoretical implications of developmental change. *Memory & Cognition*, *30*(3), 478-487. doi:10.3758/BF03194948.

Elliott, E. M., Hughes, R. W., Briganti, A., Joseph, T. N., Marsh, J. E., & Macken, B. (2016). Distraction in verbal short-term memory: Insights from developmental differences. *Journal of Memory and Language*, *88*, 39-50. doi:10.1016/j.jml.2015.12.008.

Farrell, S., Oberauer, K., Greaves, M., Pasiecznik, K., Lewandowsky, S., & Jarrold, C. (2016). A test of interference versus decay in working memory: Varying distraction within lists in a complex span task. *Journal of Memory And Language*, *90*66-87. doi:10.1016/j.jml.2016.03.010

Freeman, E., Heathcote, A., Chalmers, K., & Hockley, W. (2010). Item effects in recognition memory for words. *Journal of Memory And Language*, *62*(1), 1-18. doi:10.1016/j.jml.2009.09.004

Guitard, D., Gabel, A. J., Saint-Aubin, J., Surprenant, A. M., & Neath, I. (2018). Word length, set size, and lexical factors: Re-examining what causes the word length effect. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *44*(11), 1824-1844. doi:10.1037/xlm000055

Guitard, D., Saint-Aubin, J., Tehan, G., & Tolan, A. (2017). Does neighborhood size really cause the word length effect? *Memory & Cognition*, *46*(2), 244-260. doi:10.3758/s13421-017-0761-9

Goldhahn, D., Eckart, T., Quasthoff, U. (2012). Building large monolingual dictionaries at the Leipzig Corpora Collection: From 100 to 200 languages. *Proceedings of the Eight International Conference on Language Resources and Evaluation*.

Hanley, J. R. (1997). Does articulatory suppression remove the irrelevant speech effect? *Memory*, *5*, 423–431. doi:10.1080/741941394

Hughes, R. W. (2014). Auditory distraction: A duplex mechanism account. *PsyCh Journal*, *3*(1), 30-41. doi:10.1002/pchj.44.

Hughes, R. W., Chamberland, C., Tremblay, S., & Jones, D. M. (2016). Perceptual-motor determinants of auditory-verbal serial short-term memory. *Journal of Memory and Language*, *90,* 126-146. doi:10.1016/j.jml.2016.04.006

Hughes, R. W., Vachon, F., & Jones, D. M. (2007). Disruption of short-term memory by changing and deviant sounds: Support for a duplex-mechanism account of auditory distraction. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(6), 1050-1061. doi:10.1037/0278-7393.33.6.1050.

Hulme, C., Suprenant, A. M., Bireta, T. J., Stuart, G., & Neath, I. (2004). Abolishing the Word-Length Effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(1), 98-106. doi:10.1037/0278-7393.30.1.98

Jalbert, A., Neath, I., & Surprenant, A. M. (2011). Does length or neighborhood size cause the word length effect? *Memory & Cognition*, *39*(7), 1198-1210. doi:10.3758/s13421-011-0094-z

Jalbert, A., Neath, I., Bireta, T. J., & Surprenant, A. M. (2011). When does length cause the word length effect? *Journal of Experimental Psychology: Learning, Memory, And Cognition*, *37*(2), 338-353. doi:10.1037/a0021804

Jones, D. M., Macken, W. J., & Murray, A. C. (1993). Disruption of visual short-term memory by changing-state auditory stimuli: The role of segmentation. *Memory & Cognition*, *21*(3), 318-328. doi:10.3758/BF03208264.

Lange, E. B. (2005). Disruption of attention by irrelevant stimuli in serial recall. *Journal of Memory and Language*, *53*(4), 513-531. doi:10.1016/j.jml.2005.07.002.

Larsen, J. D., & Baddeley, A. (2003). Disruption of verbal STM by irrelevant speech, articulatory suppression, and manual tapping: Do they have a common source? The *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *56A*(8), 1249-1268. doi:10.1080/02724980244000765.

Lewandowsky, S., & Farrell, S. (2000). A redintegration account of the effects of speech rate, lexicality, and word frequency in immediate serial recall. *Psychological Research*, *63*(2), 163-173. doi:10.1007/PL00008175.

Lewandowsky, S., & Oberauer, K. (2015). Rehearsal in serial recall: An unworkable solution to the nonexistent problem of decay. *Psychological Review*, *122*(4), 674-699. doi:10.1037/a0039684

Lewandowsky, S., Oberauer, K., & Brown, G. A. (2009). No temporal decay in verbal short-term memory. *Trends in Cognitive Sciences*, *13*(3), 120-126. doi:10.1016/j.tics.2008.12.003

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*(3), 314-337. doi:10.1080/09658210344000426.

Longoni, A. M., Richardson, J. T., & Aiello, A. (1993). Articulatory rehearsal and phonological storage in working memory. *Memory & Cognition*, *21*(1), 11-22. doi:10.3758/BF03211160

Lovatt, P., Avons, S. E., & Masterson, J. (2000). The word-length effect and disyllabic words. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, *53A*(1), 1-22. doi:10.1080/027249800390646

Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing*, *19*(1), 1–36. doi:10.1097/00003446-199802000-00001

Macken, W. J., Mosdell, N., & Jones, D. M. (1999). Explaining the irrelevant-sound effect: Temporal distinctiveness or changing state? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(3), 810-814. doi:10.1037/0278-7393.25.3.810

Marian, V., Bartolotti, J., Chabal, S., Shook, A. (2012). CLEARPOND: Cross-Linguistic Easy-Access Resource for Phonological and Orthographic Neighborhood Densities. *PLoS ONE*, 7(8). doi:10.1371/journal.pone.0043230

Marsh, J. E., Hughes, R. W., & Jones, D. M. (2009). Interference by process, not content, determines semantic auditory distraction. *Cognition, 110(1*), 23–38. https://doi-org.libezp.lib.lsu.edu/10.1016/j.cognition.2008.08.003

McGill, C. I., & Elliott, E. M. (2018). Examining the role of rehearsal in auditory distraction effects in serial recall. *Manuscript In Preparation*.

Miles, C., Jones, D.M., Madden, C.A., (1991). Locus of the irrelevant sound effect in short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*, 578–584.

Murray, D. J. (1967). The role of speech responses in short-term memory. *Canadian Journal of Psychology*, *21*(3), 263-276. doi:10.1037/h0082978

Nairne, J. S. (1990). A feature model of immediate memory. *Memory & Cognition*, *18*(3), 251-269. doi:10.3758/BF03213879

Nairne, J. S. (2002). Remembering over the short term: The case against the standard model. *Annual Review of Psychology*, *53*(1), 53-81. doi:10.1146/annurev.psych.53.100901.135131

Nairne, J. S., Neath, I., & Serra, M. (1997). Proactive interference plays a role in the word-length effect. *Psychonomic Bulletin & Review*, *4*(4), 541-545. doi:10.3758/BF03214346

Neath, I. (2000). Is working memory still a useful concept? *Contemporary Psychology*, *45*(4), 410-412. doi:10.1037/002251

Neath, I., Farley, L. A., & Surprenant, A. M. (2003). Directly assessing the relationship between irrelevant speech and articulatory suppression. *Quarterly Journal of Experimental Psychology: Section A*, *56*(8), 1269-1278. doi:10.1080/02724980244000756

Rice, G. A., & Robinson, D. O. (1975). The role of bigram frequency in the perception of words and nonwords. *Memory & Cognition*, *3*(5), 513-518. doi:10.3758/BF03197523

Roodenrys, S., Hulme, C., Lethbridge, A., Hinton, M., & Nimmo, L. M. (2002). Word-frequency and phonological-neighborhood effects on verbal short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(6), 1019-1034. doi:10.1037/0278-7393.28.6.1019

Salamé, P., & Baddeley, A. D. (1982). Disruption of short-term memory by unattended speech: Implications for the structure of working memory. *Journal of Verbal Learning & Verbal Behavior*, *21*(2), 150-164. doi:10.1016/S0022-5371(82)90521-7

Schröger, E. (1997). On the detection of auditory deviations: A pre-attentive activation model. *Psychophysiology*, *34*(3), 245-257. doi:10.1111/j.1469-8986.1997.tb02395.x.

Sears, C. R., Hino, Y., & Lupker, S. J. (1995). Neighborhood size and neighborhood frequency effects in word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *21*(4), 876-900. doi:10.1037/0096-1523.21.4.876

Service, E. (1998). The effect of word length on immediate serial recall depends on phonological complexity, not articulatory duration. *Quarterly Journal of Experimental Psychology: Section A*, *51*(2), 283-304. doi:10.1080/027249898391639

Spurgeon, J., Ward, G., & Matthews, W. J. (2014). Examining the relationship between immediate serial recall and immediate free recall: Common effects of phonological loop variables but only limited evidence for the phonological loop. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(4), 1110-1141. doi:10.1037/a0035784

Vita

Corey Ian McGill, born in Plano, Texas, is a sixth-year graduate student in the Cognitive and Brain Sciences are of the Psychology Department at Louisiana State University, under the mentorship of Dr. Emily Elliott. He received a Bachelor of Science and a Master of Arts from Louisiana State University in 2013 and 2016 respectively. He has researched the role of sub-vocal speech in memory processes, the relationship between working memory and general fluid intelligence, and the effects of auditory distraction on recall. He plans to graduate with a doctorate in Psychology in May of 2019.