

UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title

Phonological Similarity Effects in Simple and Complex Word Spans

Permalink

<https://escholarship.org/uc/item/6b46q0c6>

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 28(28)

ISSN

1069-7977

Authors

Goh, Winston D.

Misra, Roopali

Publication Date

2006

Peer reviewed

Phonological Similarity Effects in Simple and Complex Word Spans

Roopali Misra (g0501312@nus.edu.sg)

Department of Psychology, National University of Singapore
11 Law Link, Singapore 117570, Singapore

Winston D. Goh (psygohw@nus.edu.sg)

Department of Psychology, National University of Singapore
11 Law Link, Singapore 117570, Singapore

Abstract

Memory span for a list of phonologically similar words is generally worse than memory span for a list of phonologically dissimilar words, a finding that is called the *phonological similarity effect*. This finding has often been cited as evidence for the use of phonological coding in short-term memory and working memory. However, some studies have demonstrated a reversal of the phonological similarity effect under certain conditions. One such condition is the use of more complex memory span tasks such as reading span. It has been suggested that sentence contexts may provide additional retrieval cues that may overcome the detrimental effects of phonological similarity. The present study examined this hypothesis by manipulating the sentence contexts of the reading span materials. No evidence showing phonological similarity facilitation was found; in fact, the standard phonological similarity decrement in recall was observed with high context sentences.

Keywords: Phonological similarity; short-term memory; working memory; serial recall; sentence context.

Introduction

One of the most well explored phenomenon in short-term memory research is the *phonological similarity effect* first reported by Conrad and Hull (1964). When a participant is presented similar and dissimilar sounding items to memorize in a serial recall task, performance is generally worse for lists of words such as *cat*, *mat*, and *hat* or letters such as *C*, *D*, *B*, and *V*, compared to lists of words or letters that do not sound similar.

However, some recent studies have found that phonological similarity does not always lead to a decrement in memory performance and in some cases, the effect reverses, i.e., performance for phonologically similar items is better than dissimilar items. Copeland and Radvansky (2001) observed that the traditional phonological similarity decrement was obtained for word span tasks where participants memorized lists of isolated words, but a phonological similarity facilitation was observed for reading span tasks in which participants read sentences and were required to recall the last word of each sentence. Other researchers have also reported reversals of the phonological similarity effect for order reconstruction tasks (Nairne & Kelley, 1999) and nonword recall (Lian, Karlsen & Eriksen, 2004). The present study will focus on the opposing effects

of phonological similarity elicited by simple versus complex span tasks as reported by Copeland and Radvansky (2001).

In simple span tasks such as digit span and word span, participants are typically presented with lists of digits or words that vary in the number (list length) of items to be memorized. The task usually begins with a short list length of 2 or 3 items and the list length is then gradually incremented in subsequent trials. Participants are required to recall the items in the order they were presented. The simple span procedure is probably the most commonly used technique to investigate serial recall performance. In complex span tasks such as reading span (Daneman & Carpenter, 1980), participants are given sentences to read and simultaneously memorize the last word of each sentence. For example, participants may read these three sentences:

Non-performers who are considered useless are fired by the boss.

While surfing on the internet, every click leads to another link.

For all humans living in the world shelter is a basic need.

They are then required to recall *boss*, *link*, and *need* at the end of the sequence. A variant of this complex span procedure is the operation span task (Turner & Engle, 1989). In this method, the sentences in the reading span task are replaced with the verification of mathematical equations such as “Is $5/10 + 3 = 2$?”. After participants verified the equation, they are then given a word to be memorized before the next equation is shown and so on. In both types of complex span, the idea is to make participants process something while trying to remember words.

Copeland and Radvansky (2001) compared the performance on phonologically similar and dissimilar words using word, reading, and operation span tasks. In word and operation span, the traditional phonological similarity effect was observed. Only the results for reading span found a reversal of the phonological similarity effect. The authors suggest that sentence contexts may provide additional retrieval cues that may overcome the detrimental effects of phonological similarity in the reading span tasks. In word and operation span tasks, the knowledge that all the words within a list rhyme may presumably provide a cue to limit

recall to words that have the same rhyme but may increase confusions about the serial position of the words within the list. This has been shown in previous studies that manipulated word similarity on semantic (Poirier & Saint-Aubin, 1995) and phonological dimensions (Crowder, 1978; Schweickert, Guentert & Hersberger, 1990). If the sentence context reverses the phonological similarity effect, then it must somehow overcome the phonological confusions that arise from similar sounding words.

One possibility is that the different sentences provide distinctive cues that allow the serial position of the processed words to be better distinguished or more salient. The additional distinctive cue provides an advantage that would not be present when memorizing isolated words without sentence contexts. In other words, in reading sentences, people may encode the sentences and use them as retrieval cues, and coupled the knowledge of the rhyming organization of the to-be-remembered words, lead to better recall.

The present study has two specific aims. The first is to determine if the phonological similarity effect reversal in the reading span task can be replicated. This is reported in Experiment 1. The second aim is determine if sentence context was in fact the reason behind the reversal. This is done in Experiment 2. Previous studies (e.g., Craik & Tulving, 1975; Schwanenflugel & Shoben, 1985) have shown that people recall words better when they are embedded in high constraint or high context sentences, compared to words embedded in low constraint or low context sentences. A more direct test of whether context is truly implicated in reversing the phonological similarity effect would be to manipulate the contextual value of the sentences in the reading span task.

If the sentence context provides an effective retrieval cue, then a high context sentence such as:

The old person drank too much wine while sitting in the bar.

would be very effective as a cue for retrieving the last word *bar* because the word is highly expected within the context of the sentence. On the other hand, a low context sentence such as:

The child wanted to celebrate his birthday so his parents took him to the bar.

would not be a very effective cue for the last word *bar*. The word is unexpected and can be easily replaced by another word.

Importantly, if Copeland and Radvansky's (2001) hypothesis that contextual retrieval cues provided by the sentence frames reverse the detrimental effects of phonological similarity, then one would expect the reversal to be greater when such cues are expected to be maximally effective in the high context sentences.

Experiment 1

The purpose of Experiment 1 is to replicate the findings of Copeland and Radvansky (2001) where the phonological similarity effect reverses in the reading span task. Participants were tested on word span and reading span tasks, using phonologically similar and dissimilar word lists. Additionally, the words recalled were scored in two different ways – the Absolute and Total span scoring methods (La Pointe & Engle, 1990). In the first method, only lists in which all words were recalled in their correct position are considered, and the Absolute Span score is the total number of words recalled in such lists. This was the same scoring procedure used by Copeland and Radvansky (2001). The absolute span method is a more conservative estimate of memory performance because it only considers trials in which all words were recalled perfectly. In the second method, the Total Span score was simply the total number of words recalled in the correct serial position in all lists, regardless of whether all the words in each list were recalled perfectly. Using two scoring procedures would determine whether the observed effects can be generalized across different measures.

To replicate the results of Copeland and Radvansky (2001), it was expected that in the word span task, recall of similar words would be worse than recall of dissimilar words. For the reading span task, recall of similar words would be better than dissimilar words.

Method

Participants Forty introductory psychology students who were native speakers of Singapore English participated for course credit.

Design and materials A 2 (Word: similar, dissimilar) x 2 (Task: word span, reading span) within-subjects design was employed.

A total of 280 words forming 40 sets of various rhymes, with each set comprising 7 words that shared the same rhyme (e.g., *bolt, colt, jolt*), were selected from the Nelson, McEvoy and Schreiber (1998) norms. The 40 sets were divided into 4 lists of 10 sets each, with each list balanced for the number of diphthongs (e.g., *bone, toy*) short (e.g., *lock, cent*), and long (e.g., *see, weed*) vowels in the rhyme sets and the types of rhymes. The lists were also equated for average log word frequency based on the Kucera and Francis (1967) counts, as revealed by a nonsignificant one-way between-subjects analysis of variance (ANOVA), $F < 1$. The average log frequency of the lists ranged from 2.24 to 2.37.

For each word, a sentence ranging from 9 to 17 words were constructed to be similar to those used by Copeland and Radvansky (2001), which ranged from 12 to 16 words. The to-be-remembered word was always at the end of the sentence.

A balanced latin-square procedure was used to rotate the lists across the 4 conditions in the study. For any one participant, a list was assigned to a single condition and was never repeated across conditions.

Procedure Participants were tested individually and all did the word span task before the reading span task. A short break was provided between each task. Within each task, half did the similar words before the dissimilar words, and vice-versa. Prior to each task, 4 practice trials at list lengths 2 and 3 using words and sentences that were not related to the experimental materials were provided.

For the word span task, words were presented on a computer monitor at a rate of 1 word per second. Participants were told to read aloud each word as they were displayed. For the reading span task, participants read aloud each sentence, and the experimenter pressed a key on the keyboard to display the next sentence. At the end of each trial, a recall prompt was shown and participants were required to verbally recall the words in the order they were presented, saying “blank” in place of words they could not recall. Responses were tape recorded.

Each condition comprised a total of 10 trials, starting with a list length of 3 and ending with a list length of 7, with 2 trials per list length. In the similar word condition, words or sentences were randomly sampled without replacement from the same rhyme set. In the dissimilar word condition, each word or sentence (up to the number required by the list length) was randomly sampled without replacement from a different rhyme set.

Results and discussion

The memory span scores for the word and reading span tasks are summarized in Table 1. A 2-way within subjects ANOVA performed on the absolute span scores revealed a significant Word x Task interaction, $F(1, 39) = 6.26$, $MSe = 10.19$, $p < .05$. The simple effects of task at both word conditions showed that word span scores were higher than reading span scores for both similar, $t(39) = 8.13$, $p < .001$, and dissimilar, $t(39) = 10.12$, $p < .001$, words. This is not surprising as reading span is the more difficult task. The more important simple effects are those between similar and dissimilar words at each span task. For word span, participants scored higher in the dissimilar condition ($M = 15.38$, $SD = 5.69$) than the similar condition ($M = 10.95$, $SD = 6.37$), $t(39) = 4.95$, $p < .001$. This shows the traditional phonological similarity effect. For reading span, participants also recalled more dissimilar ($M = 4.83$, $SD = 3.93$) than similar ($M = 2.93$, $SD = 3.36$) words, $t(39) = 2.55$, $p < .05$, again showing a phonological similarity decrement. Hence, for absolute span scores, Copeland and Radvansky’s (2001) results were not replicated. Although both span tasks show a reliable phonological similarity decrement, the significant interaction can be attributed to a smaller decrement between similar and dissimilar words in the reading span task compared to the word span task.

For the total span scores, a significant Word x Task interaction was also obtained, $F(1, 39) = 15.19$, $MSe = 16.62$, $p < .001$. Again, the simple effects of task at both word conditions showed that word span scores were higher than reading span scores for both similar, $t(39) = 3.54$, $p < .01$, and dissimilar, $t(39) = 8.32$, $p < .001$, words. As before, the important simple effects are those between

similar and dissimilar words at each span task. For word span, participants scored higher in the dissimilar condition ($M = 26.80$, $SD = 5.89$) than the similar condition ($M = 24.18$, $SD = 6.07$), $t(39) = 2.62$, $p < .05$. This shows the regular phonological similarity effect. For reading span, participants recalled more similar ($M = 19.28$, $SD = 7.45$) than dissimilar ($M = 16.88$, $SD = 6.22$) words, $t(39) = 2.38$, $p < .05$, showing a reversal of the phonological similarity effect. Hence, for the total span scores, Copeland and Radvansky’s (2001) results were replicated.

Table 1: Span scores for similar and dissimilar words across word and reading span tasks in Experiment 1.

Span score	Word span task		Reading span task	
	Similar words	Dissimilar words	Similar words	Dissimilar words
Absolute				
<i>M</i>	10.95	15.38	2.93	4.83
<i>SD</i>	6.37	5.69	3.36	3.93
Total				
<i>M</i>	24.18	26.80	19.28	16.88
<i>SD</i>	6.07	5.90	7.45	6.22

Table 2 summarizes the average number of errors across the words and tasks. We classified errors into 3 types: order errors (transpositions), omissions (“blank” responses), and item errors (extra-list intrusions). Paired samples t-tests were used to compare the differences in error types across the similar and dissimilar conditions. In general, across both word and reading span tasks, dissimilar lists elicited more omission and item errors than similar lists, but similar lists elicited more order errors.

Table 2: Mean error rates in Experiment 1.

Span task	Similar words	Dissimilar words	t-test (similar vs dissimilar)
	Order errors		
Word	13.28 (5.10)	5.23 (3.35)	$t(39) = 9.74^{**}$
Reading	9.18 (5.97)	5.28 (3.90)	$t(39) = 5.21^{**}$
	Omission errors		
Word	11.18 (5.76)	14.35 (6.67)	$t(39) = 4.04^{**}$
Reading	20.50 (6.03)	24.95 (6.84)	$t(39) = 4.76^{**}$
	Item errors		
Word	1.30 (1.60)	3.40 (2.62)	$t(39) = 5.07^{**}$
Reading	0.73 (1.13)	1.45 (1.60)	$t(39) = 3.29^*$

Note. *SDs* in parentheses; * $p < .01$, ** $p < .001$

These error patterns are consistent with previous studies manipulating word similarity on memory span (e.g., Fallon, Groves, & Tehan, 1999; Lian *et al.*, 2004; Poirier & Saint-Aubin, 1995). The similarity features could perhaps be used as global retrieval cues to help recall of the specific words in each list, but the shared features between specific words may result in positional confusion and therefore lead to more transpositions. Conversely, dissimilarity does not confer such retrieval cue advantages, and so when errors occur, they tend to be omissions or extra-list intrusions.

Different patterns of results were found in the absolute and total span scores for reading span, where the absolute span scores showed the traditional phonological similarity decrement while the total span scores supported Copeland and Radvansky's phonological similarity facilitation. As the absolute span scoring procedure has a stricter criterion in which all items at any particular list length must be recalled in perfect order before they are counted, an examination of performance at each list length may shed some light on precisely which list length may be responsible for the disparate results. It is possible that performance indices may be obscured at the shorter (e.g., list length 3) and the largest (e.g., list length 7) list lengths due to differences in cognitive load. The best indicator of the direction of phonological similarity effects may be found at list lengths which approach the average short-term memory capacity for words of participants. Figure 1 shows the proportion of correctly recalled words across the various list lengths.

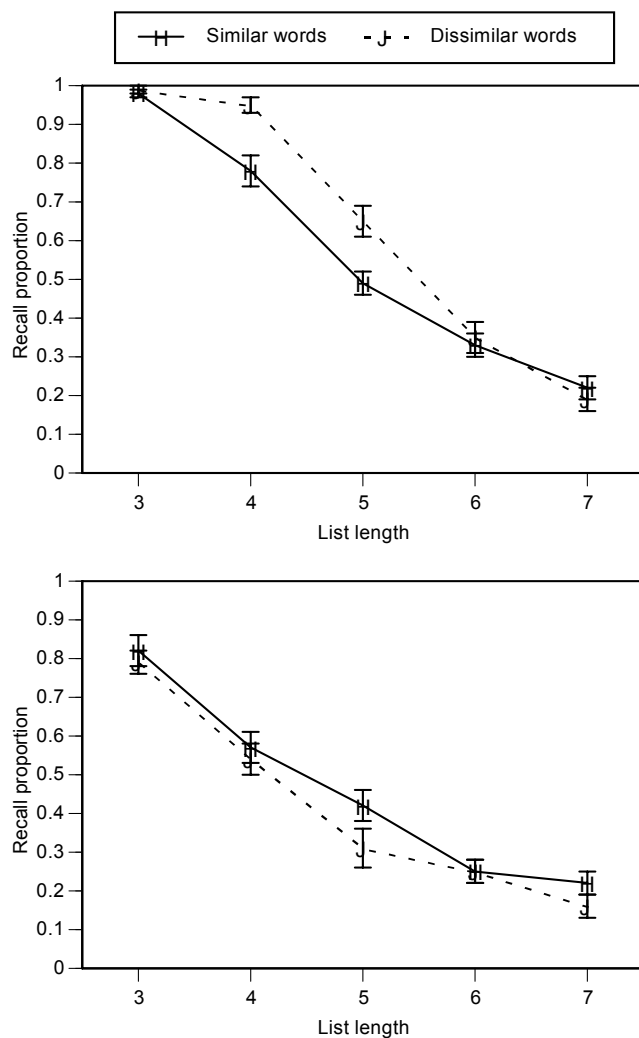


Figure 1: Average recall probability across trials (+ SEs) as a function of word similarity in the word span (top panel) and reading span (bottom panel) tasks.

A 3-way within-subjects ANOVA revealed a significant Word x Task x Length interaction, $F(4, 156) = 6.38$, $MSe =$

0.02, $p < .001$. To determine the source of the interaction, the simple effects of word at each list length in the two span tasks were examined. For the word span task, the similar-dissimilar difference at list lengths 3, 6, and 7 were not significant, all $t(39) < 1.06$. At list length 4, more dissimilar words ($M = .95$, $SD = .10$) were recalled than similar words ($M = .78$, $SD = .23$), $t(39) = 4.34$, $p < .001$. The same pattern was found at list length 5, $t(39) = 3.48$, $p < .001$, more dissimilar words ($M = .65$, $SD = .25$) were recalled than the similar words ($M = .49$, $SD = .21$). For the reading span task, the similar-dissimilar difference was found to be significant at list length 5 only, $t(39) = 2.92$, $p < .01$, showing a phonological similarity facilitation with more similar words ($M = .42$, $SD = .24$) correctly recalled than dissimilar words ($M = .31$, $SD = .23$). The differences at all other list lengths were not significant, all $t(39) < 1.81$.

The list length analyses show patterns supporting Copeland and Radvansky (2001), but it occurs only at list length 5 where it may be at the limit of short-term memory capacity for most participants. There may be ceiling and floor effects at the shorter and longer list lengths preventing phonological similarity effects from emerging. The next experiment will attempt to directly test the notion that sentence context is the reason behind the phonological similarity facilitation found in the reading span task.

Experiment 2

If sentence context provides information that will overcome the usual phonological confusions arising from memorizing similar sounding words, it may be possible to enhance or reduce this effect by manipulating the contextual value of the sentences. The present experiment examined reading span performance for similar and dissimilar words using high and low context sentences. If the hypothesis that sentence context does provide information that will overcome the detrimental effects of phonological similarity, one would expect the phonological similarity reversal to be greatest with high context sentences and attenuated or eliminated with low context sentences. In light of the findings from Experiment 1 that most of the differences between similar and dissimilar words occurred at about list length 5, a fixed length procedure using list length 5 was employed in the present experiment to avoid potential obscuring of the results by ceiling and floor effects.

Method

Participants Forty introductory psychology students who did not take part in the previous experiment participated for course credit. All were native speakers of Singapore English.

Design and materials A 2 (Word: similar, dissimilar) x 2 (Context: high, low) within-subjects design was employed.

The fixed length procedure required fewer words than the memory span procedure used in Experiment 1. 100 words from the 280-word pool used in the previous experiment were selected and balanced for the same attributes. The words formed 20 sets of rhymes with each set comprising 5 rhyming words. The 20 sets were divided into 4 lists of 5

sets each, with the average log frequency, ranging from 2.44 to 2.52, balanced across lists, $F < 1$.

For each word, one high and one low context sentence were constructed with the same structure as the sentences in Experiment 1, in which the to-be-remembered word was always at the end of the sentence. To ensure that the high and low context sentences were appropriate, 40 participants who did not take part in either experiment rated the predictability of the last word for each sentence. Inappropriate sentences were reconstructed and checked again with additional participants.

A balanced latin-square procedure similar to the one used in Experiment 1 was adopted to rotate the lists across the 4 conditions.

Procedure The procedure was similar to the reading span task in Experiment 1, except that each condition comprised a total of 5 trials of 5 sentences each, following the fixed list length procedure. The order in which the 4 conditions were presented was counterbalanced across participants. As the memory span procedure using variable list lengths was not used in this experiment, absolute and total span scores were not relevant. Scoring was based on the proportion of correct-in-position serial recall.

Results and discussion

Serial recall performance is summarized in Table 3. A two-way within-subjects ANOVA revealed a significant Word x Context interaction, $F(1, 39) = 17.10$, $MSe = 0.01$, $p < .001$. Tests of simple effects showed that with high context sentences, recall of dissimilar words ($M = .57$, $SD = .12$) was higher than similar words ($M = .49$, $SD = .14$), $t(39) = 4.54$, $p < .001$. On the other hand, for low context sentences, no differences in recall was found between similar ($M = .45$, $SD = .17$) and dissimilar words ($M = .42$, $SD = .14$), $t(39) = 1.09$, *ns*. Thus, the hypothesis that the reversal of the phonological similarity effect is due to sentence context is not supported. In fact, it appears that a traditional phonological similarity decrement was observed with the high context sentences, but eliminated with low context sentences.

Table 3: Recall proportions for similar and dissimilar words in high and low context sentences in Experiment 2.

Recall Proportion	High Context		Low Context	
	Similar	Dissimilar	Similar	Dissimilar
<i>M</i>	.49	.57	.45	.42
<i>SD</i>	.14	.12	.17	.14

In the similar condition, recall of words in high context sentences ($M = .49$, $SD = .14$) was greater than words in low context sentences ($M = .45$, $SD = .17$), $t(39) = 2.30$, $p < .05$. The same pattern was observed in the dissimilar condition, recall for words in high context sentences ($M = .57$, $SD = .12$) was greater than words in low context sentences ($M = .42$, $SD = .14$), $t(39) = 7.68$, $p < .001$. This result is consistent with previous findings showing context

availability effects on memory (e.g. Schwanenflugel & Shoben, 1983; Wattenmaker & Shoben, 1987).

Table 4 summarizes the error patterns. In general, paired samples t-tests revealed essentially the same results found in Experiment 1. Except for the lack of any reliable difference in omissions between similar and dissimilar words in high context sentences, there were more order errors for similar words and more item errors and omissions for dissimilar words.

Table 4: Mean error rates in Experiment 2.

Context	Similar words	Dissimilar words	t-test (similar vs dissimilar)
Order errors			
High	3.68 (2.46)	1.10 (1.66)	$t(39) = 7.85^{**}$
Low	4.28 (2.96)	1.68 (1.91)	$t(39) = 6.72^{**}$
Omission errors			
High	9.05 (3.54)	8.70 (3.20)	$t(39) = 0.69$
Low	8.73 (4.16)	11.18 (3.62)	$t(39) = 4.60^{**}$
Item errors			
High	0.20 (0.46)	0.58 (0.93)	$t(39) = 2.07^*$
Low	0.33 (0.53)	1.30 (1.45)	$t(39) = 4.78^{**}$

Note. *SDs* in parentheses; * $p < .05$, ** $p < .001$

General Discussion

In Experiment 1, the pattern of results in the word span task was consistent with the traditional phonological similarity effect; dissimilar words were better recalled than similar words. There was some evidence that this effect reversed in the reading span task when total span scores and list length analyses were considered, replicating the trend reported by Copeland and Radvansky (2001). However, with absolute span scores, the effect showed a phonological similarity decrement rather than facilitation. These inconsistent findings with reading span suggest that it may be important to use multiple scoring procedures to probe differential performance in future studies so as to ascertain the generalizability of the effects across measures.

As most of the action in Experiment 1 appeared to be taking place at about list length 5, Experiment 2 was conceived to directly test Copeland and Radvansky's (2001) argument that the reversal of the phonological similarity effect is dependent on whether contextual information can overcome or attenuate phonological confusions. It was hypothesized that recall for similar words would be better than dissimilar words for sentences with high contextual value. This reversal did not occur; instead the traditional phonological similarity effect appears to be present in the high context sentences and eliminated in the low context sentences.

Although the reading span results across the two experiments were not entirely consistent, there are two reasons why we think a conclusion that sentence context does not cause a reversal of the phonological similarity effect is viable. First, the overall pattern in Experiment 2 showed that context availability led to general recall facilitation, which was evident from the recall scores in the high and low context sentences in both similar and

dissimilar conditions. Recall was better when participants were presented high context sentences than when they were provided with low context sentences. These results are consistent with previous studies on context effects on memory (e.g., Craik & Tulving, 1975; Schwanenflugel & Shoben, 1983; Schwanenflugel & Shoben, 1985; Wattenmaker & Shoben, 1987). Craik and Tulving (1975) emphasizes the fact that one of the possible ways in which memory codes could be elaborated would be to provide a richer and better context. However, this manipulation of context clearly did not diminish the phonological similarity decrement in the expected direction. It may have enhanced the availability of item information in high context sentences so that participants could remember the last words better (which is consistent with better overall recall in high context sentences), but if the words all sounded alike, then confusions as to the serial position of these words ensued, leading to the traditional phonological similarity effect. In low context sentences, item information suffers because the sentence frame does not predict the last word, and so there is a general decrement in recall regardless of whether the words sound similar or not, leading to an elimination of phonological similarity effects.

Second, this interpretation is supported by the error patterns, which were remarkably consistent across the two experiments. More order errors were observed in the similar conditions, and more omissions and item errors were observed in the dissimilar conditions. These results replicate and extend previous findings that show dissociations between item and order information when similarity is manipulated (e.g. Fallon *et al.*, 1999; Lian *et al.*, 2004; Poirier & Saint-Aubin, 1995).

In summary, the main goal of the present study was to examine Copeland and Radvansky's (2001) hypothesis that sentence context in the reading span test was responsible for the phonological similarity reversal in their experiments. In this regard, there was no strong evidence that the phonological similarity effect would be reversed when the contextual value of the sentences is manipulated.

Acknowledgments

This work was supported in part by a National University of Singapore research scholarship awarded to RM and Research Grant R-581-000-048-112 to WDG. We thank David Copeland for samples of the materials used in his study.

References

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

Copeland, D. E., & Radvansky, G. A. (2001). Phonological similarity in working memory. *Memory & Cognition*, *29*, 774-776.

Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General*, *104*, 268-294.

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

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, *19*, 450-466.

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

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

La Pointe, L. B., & Engle, R. W. (1990). Simple and complex word span as measures of working memory capacity. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *16*, 1118-1130.

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

Nairne, J. S., & Kelley, M. R. (1999). Reversing the phonological similarity effect. *Memory & Cognition*, *27*, 45-53.

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1998). The University of South Florida word association, rhyme, and word fragment norms. <http://www.usf.edu/FreeAssociation/>.

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

Schwanenflugel, P. J., & Shoben, E. J. (1983). Differential context effects in the comprehension of abstract and concrete verbal materials. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *9*, 82-102.

Schwanenflugel, P. J., & Shoben, E. J. (1985). The influence of sentence constraint on the scope of facilitation for upcoming words. *Journal of Memory and Language*, *24*, 232-252.

Schweickert, R., Geunert, L., & Hersberger, L. (1990). Phonological similarity, pronunciation rate, and memory span. *Psychological Science*, *1*, 74-77.

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

Wattenmaker, W. D., & Shoben, E. J. (1987). Context and the recallability of concrete and abstract sentences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13*, 140-150.