

Phonological similarity effects in simple and complex span tasks

Brooke N. Macnamara · Adam B. Moore ·
Andrew R. A. Conway

© Psychonomic Society, Inc. 2011

Abstract Three experiments were conducted to examine the effect of phonological similarity in simple and complex memory span tasks. In Experiment 1, participants performed either a simple or a complex span task, and the memoranda within lists were either phonologically similar or distinct. Phonologically similar lists consisted of words that rhymed. The simple span task was word span. There were two complex span tasks; one was the original reading span task, and the other was a variant of reading span in which all the sentences within a list were contextually related. The classic phonological similarity decrement was observed in word span. In contrast, phonological similarity facilitation was observed in both versions of reading span. This facilitation effect was further investigated in Experiment 2 using two new versions of reading span. In Experiment 2, the sentences in reading span were either short or long, and the memoranda were presented separately from, and were unrelated to, the sentences. Again, words within phonologically similar lists rhymed, and again, facilitation was observed. In Experiment 3, phonological similarity was operationalized in terms of feature overlap, rather than rhyme. The classic phonological similarity decrement was still observed in word span, but facilitation was not observed in complex span. The results suggest that phonological similarity, when operationalized using words that rhyme, serves as a list retrieval cue and that complex span tasks are more dependent on cue-driven memory retrieval mechanisms than are simple span tasks.

Keywords Span task · Working memory ·
Phonological similarity

B. N. Macnamara · A. B. Moore · A. R. A. Conway (✉)
Department of Psychology, Princeton University,
Green Hall,
Princeton, NJ 08540, USA
e-mail: aconway@princeton.edu

In the present experiments, we investigated the effect of phonological similarity on simple and complex span performance. The phonological similarity effect, which refers to the finding that phonologically dissimilar lists of stimuli are recalled better than phonologically similar lists, is considered a classic finding in short-term memory research and has been reliably demonstrated in simple span tasks and other tasks that require immediate serial recall (for reviews, see Botvinick & Plaut, 2006; Lewandowsky & Farrell, 2008; Page & Norris, 1998). However, phonological similarity has been investigated less in complex span, which is somewhat surprising because it does not always have an effect on complex span and, in some cases, the effect is reversed, resulting in phonological similarity facilitation (Copeland & Radvansky, 2001; Tehan, Hendry, & Kocinski, 2001).

To be clear, simple span tasks consist of a list of to-be-remembered stimuli, presented one at a time—typically, at a rate of one stimulus per second—followed by a recall prompt, at which point the participant is required to recall the stimuli in the serial order in which they were presented. In contrast, complex span tasks require the participant to perform some secondary task while encoding and maintaining stimuli for later recall. For example, in the reading span task, the participant must read a series of sentences, each presented one at a time, and remember the last word of each sentence for later recall. As in simple span, participants are typically required to recall the stimuli in the serial order in which they were presented.

The difference between simple and complex span tasks is of theoretical interest because these tasks have shown empirical dissociations in both psychometric and neuroimaging investigations of working memory (WM). Psychometrically, simple span tends to be less predictive than complex span of cognitive ability and general fluid intelligence (Conway, Cowan, Bunting, Theriault, &

Minkoff, 2002; Conway, Kane, & Engle, 2003; Daneman & Carpenter, 1980; Daneman & Merikle, 1996; Engle, Tuholski, Laughlin, & Conway, 1999; Kane et al., 2004). Also, cross-domain correlations are higher for complex span than for simple span tasks. That is, correlations among verbal and spatial complex span tasks tend to be higher than correlations among verbal and spatial simple span tasks (Kane et al., 2004). Finally, verbal simple span predicts verbal reasoning more strongly than it predicts spatial reasoning, and likewise, spatial simple span predicts spatial reasoning more strongly than it predicts verbal reasoning (Kane et al., 2004). These results collectively suggest that complex span tasks are more dependent upon domain-general cognitive mechanisms than are simple span tasks.

Neuroimaging experiments also suggest important distinctions between simple and complex span tasks (Bunge, Klingberg, Jacobsen, & Gabrieli, 2000; Kondo et al., 2004; M. Osaka et al., 2003; N. Osaka et al., 2004; Smith et al., 2001). Of particular interest here are experiments on complex span, because they typically include simple span as a baseline (contrast) condition. For example, Bunge et al. had participants perform reading span while undergoing fMRI. The participants also completed two baseline conditions, a reading-only task and a remember-words-only task. Neural activity during the reading span task was contrasted with that during the reading-only task and, separately, with that during the remember-words-only task, and then the conjunction of the two contrasts was taken to identify brain regions uniquely recruited by reading span. This analysis revealed neural activity in the bilateral prefrontal cortex and left parietal cortex. Several other fMRI experiments have subsequently replicated this result, and lateral prefrontal and posterior parietal areas are commonly thought to contribute to WM function (for a review, see Jonides et al., 2008).

One recent fMRI experiment on complex span is particularly relevant to the present investigation. Chein, Moore, and Conway (2011) had participants perform both verbal and spatial complex span and their corresponding baseline conditions. The results replicated previous fMRI studies on complex span, such that the lateral prefrontal cortex and posterior parietal cortex were more active in complex span than in the baseline conditions, as revealed by conjunction analyses like the one described above. More important, when fMRI data acquisition was restricted to the recall stage of a trial (rather than averaging across the entire trial, as had been done in previous research), there was greater activity in the hippocampus during complex span than during simple span. This suggests that complex span tasks recruit memory retrieval mechanisms that are typically associated with retrieval from long-term memory.

These recent fMRI results are consistent with a theoretical account of individual differences in WM

capacity (WMC) proposed by Unsworth and Engle (2007). According to their view, individual differences in WMC arise from two abilities: the ability to maintain information in an immediately accessible state and the ability to retrieve information that is not immediately accessible. In their terms, WM engages both primary and secondary memory. They argued that primary memory is responsible for maintaining a limited number of highly active representations in short-term intervals. The representations are maintained through continuous allocation of attention so they are immediately accessible. Primary memory is generally assumed to be limited to approximately four representations (Atkinson & Shiffrin, 1968; Cowan, 2001). According to Unsworth and Engle, when a task requires that more than four items be maintained simultaneously, when attention is diverted toward processing new material, or when attention is captured by other stimuli, representations are displaced from primary memory into secondary memory. These representations must then be recovered via a cue-dependent search mechanism (Unsworth & Engle, 2007).

Thus, according to this view, in complex span tasks, the to-be-remembered items are quickly displaced into secondary memory, due to the processing component of the task (e.g., reading sentences aloud in reading span), and must therefore be retrieved at recall via a discriminative search process. This could explain why phonological similarity has a different effect on simple and complex span. Complex span is more dependent on memory retrieval mechanisms associated with the medial temporal lobe (Chein et al., 2011). Thus, the fact that the list of items is phonologically similar may serve as a retrieval cue and facilitate recall. An alternative idea, which is not mutually exclusive, is that the nature of memory representations in complex span tasks change as a function of time and/or interference and, as a result, phonological features become less salient. This notion will be further considered in the General Discussion section.

Manipulating phonological similarity in complex span is not a novel experiment. For example, Copeland and Radvansky (2001) demonstrated the classic phonological similarity *decrement* in a simple span task, in which the to-be-remembered items were words, and phonological similarity *facilitation* in a reading span task, in which the to-be-remembered items were the final word of each sentence (see also Tehan et al., 2001). However, Copeland and Radvansky found no effect of phonological similarity when the interpolated task in complex span was solving math problems (i.e., the operation span task; Turner & Engle, 1989). On the basis of these findings, Copeland and Radvansky concluded that enhanced recall in the phonologically similar condition in reading span was due to knowledge that the target words rhymed coupled with the

context of the sentences. This view is therefore referred to here as the *sentence context hypothesis*.

Nairne and Kelley (1999) were also able to reverse the phonological similarity effect by employing a Brown–Peterson type task (Brown, 1958; Peterson & Peterson, 1959). In Brown–Peterson tasks, a list of to-be-remembered items is presented, and then a distracting task is performed prior to recall. Nairne and Kelley had participants read digits aloud after a list of stimuli had been presented. In one condition, participants read only 4 digits, and in another condition, they read 48 digits. In the former condition, the classic phonological similarity decrement was observed, but in the latter condition, phonological similarity facilitation was observed. Thus, Nairne and Kelly produced a phonological similarity facilitation effect in a task that did not include sentence context. It is therefore possible that phonological similarity itself serves as a list retrieval cue (Nairne & Kelly, 1999) and this cue is what causes phonological similarity facilitation in reading span. We therefore refer to this view as the *list context hypothesis*.

The first two experiments reported here tested whether phonological similarity facilitation in reading span is caused by sentence context or simply because similarity is an efficient list retrieval cue. In the first experiment, we manipulated phonological similarity in word span and two versions of reading span, in which we manipulated sentence context. In the second experiment, we again used reading span tasks but separated the to-be-remembered words from the sentences, thus eliminating the link between sentence context and the memoranda. To preview, phonological similarity facilitation was observed in all reading span conditions across the two experiments, supporting the list context hypothesis.

This facilitation effect was further explored in Experiment 3. In our first two experiments and in Copeland and Radvansky (2001), phonologically similar lists consisted of words that rhyme. An alternative approach is to create lists consisting of words that share phonological features but do not rhyme. We adopted this approach in Experiment 3 to demonstrate a boundary condition on the facilitation effect. That is, we did not expect to observe facilitation in this case, because feature overlap is not as salient as rhyme and, therefore, should not serve as an efficient retrieval cue.

Experiment 1

Method

Participants

Sixty-one undergraduate students participated in Experiment 1 (E1) in exchange for partial course credit. Participants were

randomly assigned to either a simple span task or a complex span task. Participants assigned to the simple span task completed word span ($n = 20$). Participants assigned to a complex span task completed either reading span ($n = 20$) or story span ($n = 21$). The participants were all native English speakers.

Materials and procedure

A total of 108 single-syllable nouns were used as memoranda. The words were normed for frequency (Kučera & Francis, 1967), meaningfulness (Toglia & Battig, 1978), familiarity (pooled; Gilhooly & Logie, 1980; Pavio, 2011; Toglia & Battig, 1978), concreteness (pooled; Gilhooly & Logie, 1980; Pavio, 2011; Toglia & Battig, 1978), and imageability (pooled; Gilhooly & Logie, 1980; Pavio, 2011; Toglia & Battig, 1978), using the MRC Psycholinguistic Database (Coltheart, 1981).

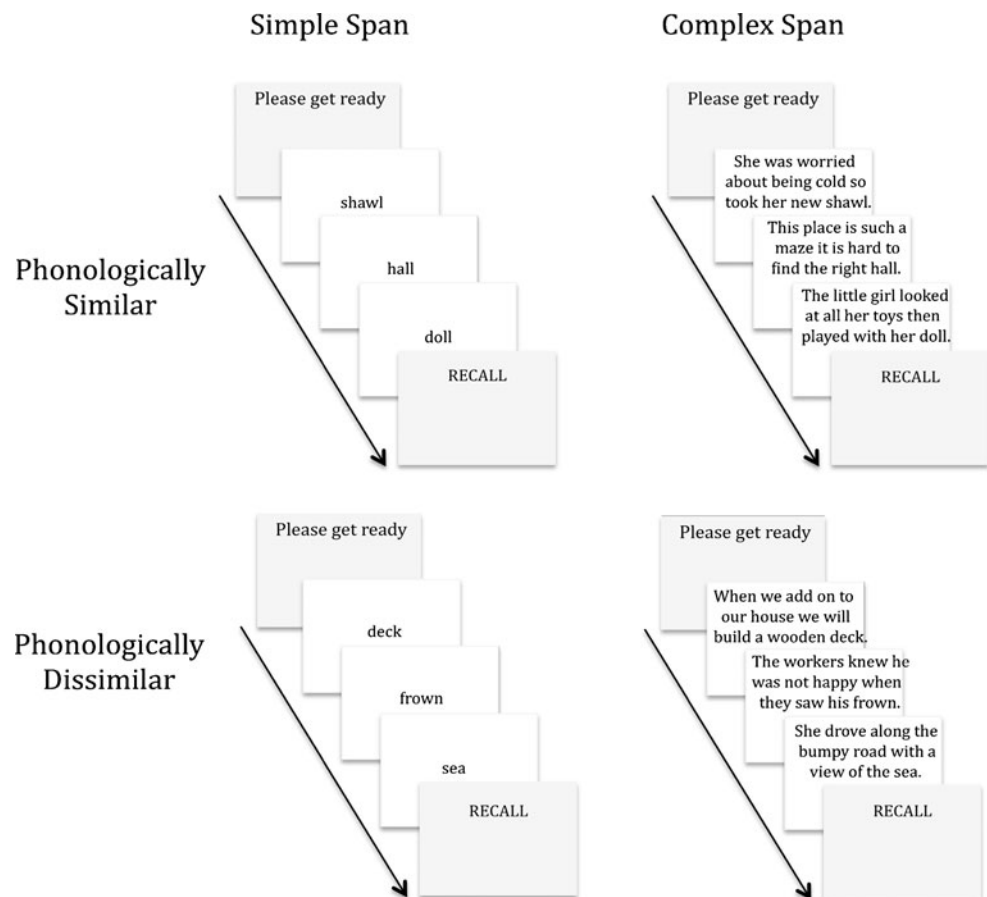
Fifty-four of the words were arranged in three lists of 3, 4, 5, and 6 such that the words within each list were phonologically similar to one another (e.g., *shawl*, *hall*, *doll*). The other 54 words were arranged in three lists of 3, 4, 5, and 6 such that the words within each set were phonologically dissimilar from one another (e.g., *deck*, *frown*, *sea*). All tasks were experimenter paced.

Word span Words were presented one at a time on a computer screen, and participants were instructed to read the words aloud. At the end of a list, the word “RECALL” was presented, and participants were instructed to type all the words into a computer in correct serial order. After participants practiced the task with four 2-length lists, two phonologically similar and two dissimilar, they began the actual task.

The task consisted of three 3-length lists, three 4-length lists, three 5-length lists, and three 6-length lists in each condition. Phonological similarity was manipulated within participants and blocked, and the blocks were counter-balanced across participants, such that half the participants performed the phonologically similar condition first and half performed the dissimilar condition first. The lists within each phonological condition were presented randomly, such that the participant could not predict the length of the upcoming list. Prior to the presentation of each new list, a ready screen appeared for 2,000 ms (see the left column of Fig. 1).

Reading span One hundred eight sentences were created, containing from 14 to 19 syllables. Sentences were normed for reading on the basis of combined scores from the Flesch Reading Ease Formula (Flesch, 1948), pooled Flesch–Kincaid, (Kincaid, Fishburne, Rogers, & Chissom, 1975), Coleman–Liau (Coleman & Liu, 1975), Gunning Fog

Fig. 1 Examples of phonologically similar and phonologically dissimilar length-three lists in a simple span task (word span) and in a complex span task (reading span)



(Gunning, 1952), and SMOG (McLaughlin, 1969) indices. Each sentence was created so that it ended with one of the 108 single-syllable nouns used as stimuli in the word span task.

In the complex span tasks, sentences were presented one at a time on a computer screen, and participants were instructed to read the sentences aloud and attempt to remember the final word in each sentence. Half the final words of the sentences were phonologically similar within a list of sentences, and half were dissimilar. For example, participants read this 3-length list in the phonologically dissimilar condition: (1) "When we add on to our house we will build a wooden deck"; (2) "The workers knew he was not happy when they saw his frown"; (3) "She drove along the bumpy road with a view of the sea." The experimenter clicked the mouse to present the next sentence immediately after the participant had finished reading the previous sentence, in an effort to reduce rehearsal. After participants practiced the task with four 2-length lists, two phonologically similar and two dissimilar, they began the actual task.

The tasks consisted of three 3-length lists, three 4-length lists, three 5-length lists, and three 6-length lists in each condition. Phonological similarity was manipulated within

participants and blocked, and the blocks were counter-balanced across participants, such that half the participants performed the phonologically similar condition first and half performed the dissimilar condition first. The lists within each phonological condition were presented randomly, such that the participant could not predict the length of the upcoming list. Prior to the presentation of each new list, a ready screen appeared for 2,000 ms (see the right column of Fig. 1).

Story span A second version of reading span was created, which we refer to as *story span*. The only difference between reading span and story span was that, in story span, the sentences were contextually related within each list. For example, a participant in the reading span condition saw this three-length phonologically dissimilar list in which the sentences *are not* contextually related:

1. "When we add on to our house we will build a wooden deck."
2. "The workers knew he was not happy when they saw his frown."
3. "She drove along the bumpy road with a view of the sea."

A participant in the story span condition saw this three-length phonologically dissimilar list in which the sentences *are* contextually related:

1. “The captain needed to talk to his crew so called for all hands on deck.”
2. “The sailors knew something was very wrong when they saw his frown.”
3. “There was an unavoidable storm and they were out on the open sea.”

Scoring

Phonological similarity effects are most often observed in serial recall. However, other scores can be derived from the data and have the potential to clarify the nature of similarity effects (Neale & Tehan, 2007). We therefore used three scoring procedures, resulting in three dependent measures, which we refer to as *serial recall*, *item recall*, and *order accuracy*. Each of these measures is described in more detail below. Detailed error analyses were also conducted for each experiment, but for space considerations, they are reported in a table in the [Appendix](#).

Serial recall The partial-credit-load method was used to score all the span tasks (Conway et al., 2005). Partial-credit methods consider a list partially correct if some of the items were recalled correctly. For example, if two out of three words were recalled in correct serial order, credit would be awarded for the two correctly recalled words (for serial recall scoring). Load methods of scoring take list length into account. For example, if three out of three words were recalled in correct serial order, the score would be 3. If six out of six words were recalled in correct serial order, the score would be 6. List scores are then added and divided by the total number of memoranda—in this case, 54 in each condition. Thus, the dependent measure is the proportion of items recalled in the correct serial position.

Item recall The partial-credit-load method was also used for item recall. However, for item recall, an item was considered correct if it was recalled in any serial position. Lists were scored and summed in the same manner as described above, and the dependent measure was the proportion of items recalled, regardless of serial position.

Order accuracy Order accuracy refers to the ratio of serial recall to item recall. In other words, if an item is recalled, what is the probability that it is recalled in the correct serial position? For each list we calculated the ratio of items recalled in correct serial order, relative to total items recalled. For example, if a participant recalled two items from a five-item list but only one of the two

items was in the correct serial position, order accuracy would be .5.

Results and discussion

All statistical significance tests were conducted from the perspective of null hypothesis significant testing with $\alpha = .05$, and effect sizes were estimated using eta-squared, η^2 , or partial eta-squared, η_p^2 .

Serial recall A 3×2 mixed factorial ANOVA revealed a significant main effect of task (word span, reading span, or story span), $F(2, 58) = 10.54, p < .01, \eta_p^2 = .27$. The main effect of similarity (phonologically similar or dissimilar) was not significant, $F(1, 58) = 2.18, p = .15, \eta_p^2 = .04$, due to the opposing effects of phonological similarity on simple and complex span tasks, as revealed by a significant interaction, $F(2, 58) = 22.62, p < .01, \eta_p^2 = .44$ (see Fig. 2a).

There was considerably more variance in the complex span tasks than in the simple span task, violating the homogeneity-of-variance assumption of the ANOVA. Indeed, Levene’s test was significant for both the similar and dissimilar conditions (for both, $p < .05$). Therefore, the significant interaction from the omnibus analysis was further examined by conducting two simple effects analyses, one on the simple span data and another on the complex span data.

The simple effects analysis on word span revealed a significant main effect of phonological similarity, such that serial recall was greater in the dissimilar condition than in the similar condition, $F(1, 19) = 78.8, p < .01, \eta^2 = .81$, replicating the classic phonological similarity effect.

Simple effects analyses on the complex span data consisted of a 2×2 mixed factorial ANOVA with task (reading span, story span) and phonological similarity (similar, dissimilar) as the independent variables. This analysis revealed phonological similarity *facilitation*; that is, there was a significant main effect of phonological similarity, such that serial recall was greater in the similar condition than in the dissimilar condition, $F(1, 39) = 5.89, p = .02, \eta_p^2 = .13$. The main effect of task was not significant, $F(1, 39) = 2.58, p = .12, \eta_p^2 = .06$, but there was a trend in both the similar and dissimilar conditions indicating that serial recall was greater in story span than in reading span (see Fig. 2a), suggesting that the added context in story span had an effect on serial recall, albeit a weak one. More important, the interaction between task and similarity was not significant (and in fact, it was completely nonexistent), $F(1, 39) = 0.00, p = .99, \eta_p^2 = .00$. Taken together, these results suggest that while sentence (or story) context slightly improved serial recall, it did not impact phonological similarity facilitation, which is incon-

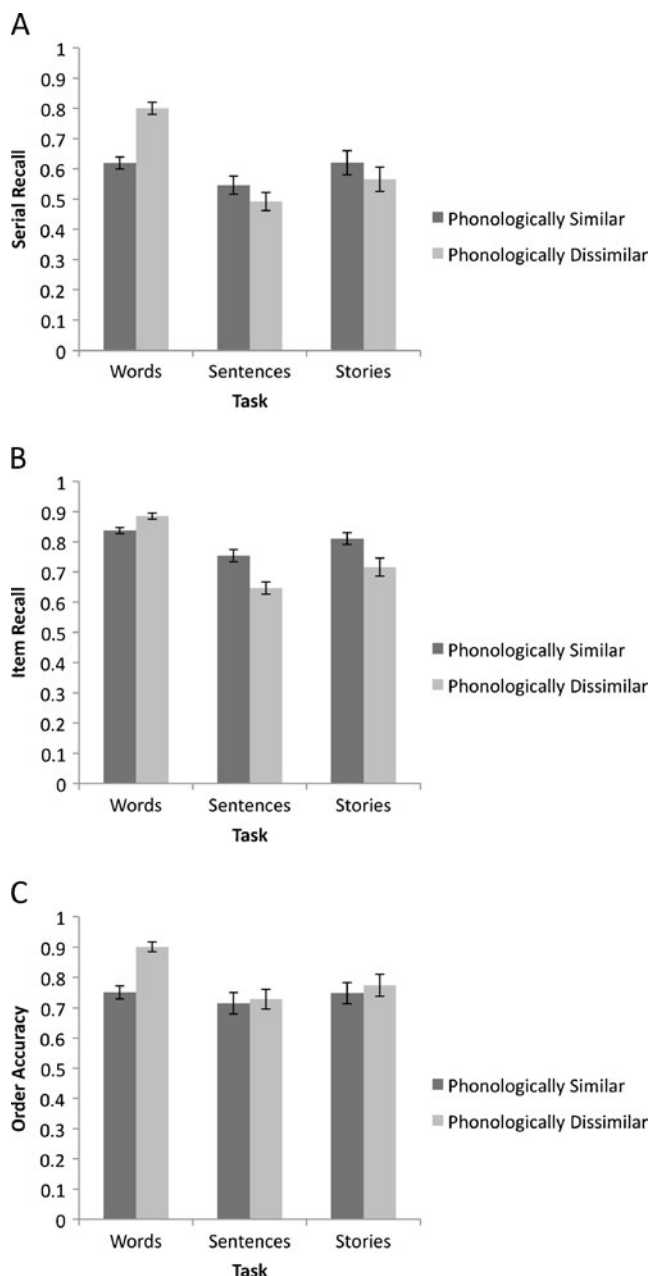


Fig. 2 Serial recall, item recall, and order accuracy in Experiment 1. Words within phonologically similar lists rhymed. Error bars represent the standard error of the mean

sistent with the sentence context hypothesis. Instead, it seems that phonological similarity itself serves as a retrieval cue, facilitating recall in complex span.

Item recall A 3×2 mixed factorial ANOVA was also conducted on item recall. This analysis revealed a significant main effect of task, $F(2, 58) = 23.15$, $p < .01$, $\eta_p^2 = .44$, and a significant main effect of phonological similarity, $F(1, 58) = 27.67$, $p < .01$, $\eta_p^2 = .32$. However, these main effects were qualified by a

significant interaction between task and similarity, $F(2, 58) = 25.78$, $p < .01$, $\eta_p^2 = .47$ (see Fig. 2b). Again, there was significantly more error variance in the complex span tasks than in the word span task (Levene's test was significant in both the dissimilar and similar conditions; $p < .05$ for both), so the interaction was further examined by considering the simple and complex span data separately.

For the word span task, the classic phonological similarity effect was again observed, $F(1, 19) = 11.07$, $p < .01$, $\eta_p^2 = .37$, such that item recall was better for dissimilar lists than for similar lists. To further examine the complex span data, a 2×2 mixed factorial ANOVA was conducted with task as the between-groups variable and similarity as the within-groups variable. Here, we observed phonological similarity facilitation, $F(1, 39) = 63.15$, $p < .00$, $\eta_p^2 = .62$, such that item recall was better for similar lists than for dissimilar lists. The main effect of task was also significant, $F(1, 39) = 5.57$, $p = .02$, $\eta_p^2 = .13$, such that item recall was better in story span than in reading span. The interaction was not significant, $F(1, 39) = .21$, $p = .65$, $\eta_p^2 = .01$. These results are consistent with the serial recall data, except that, here, the difference between story span and reading span reached significance.

Order accuracy A 3×2 mixed factorial ANOVA was also conducted on order accuracy. This analysis revealed a significant main effect of task, $F(2, 58) = 3.54$, $p = .04$, $\eta_p^2 = .11$, and a significant main effect of phonological similarity, $F(1, 58) = 16.68$, $p < .01$, $\eta_p^2 = .22$. However, these main effects were qualified by a significant interaction between task and similarity, $F(2, 58) = 22.62$, $p < .01$, $\eta_p^2 = .44$ (see Fig. 2c). Again, there was significantly more error variance in the complex span tasks than in the word span task (Levene's test was significant in the dissimilar condition, $p = .02$, and close to significant in the similar condition, $p = .11$), so the interaction was further examined by considering the simple and complex span data separately.

For the word span task, the classic phonological similarity effect was again observed, $F(1, 19) = 83.18$, $p < .01$, $\eta_p^2 = .81$, such that order accuracy was better for dissimilar lists than for similar lists. To further examine the complex span data, a 2×2 mixed factorial ANOVA was conducted with task as the between-groups variable and similarity as the within-groups variable. None of the effects in this analysis were significant [task, $F(1, 39) = 0.79$, $p = .38$, $\eta_p^2 = .02$; similarity, $F(1, 39) = 0.84$, $p = .37$, $\eta_p^2 = .02$; interaction, $F(1, 39) = 0.08$, $p = .78$, $\eta_p^2 = .00$]. The significant interaction in the omnibus analysis is therefore due to the fact that phonological similarity has an effect on order accuracy in simple span, but not complex span.

To summarize the results of E1, the classic phonological similarity decrement was observed in simple span and in all three dependent variables. In contrast, phonological similarity facilitation was observed in both versions of reading span, in serial recall and item recall. Also, serial recall and item recall were better in story span than in reading span, suggesting that the added context in story span did improve recall. However, the facilitation effect was the same in story span and reading span, suggesting that sentence context does not have an effect on phonological similarity facilitation. This was further tested in Experiment 2.

Experiment 2

Experiment 2 (E2) was a 2×2 mixed factorial design in which participants performed one of two variations of reading span while phonological similarity was manipulated within the tasks in the same manner as in E1. However, instead of reading sentences in which the final word was the to-be-remembered item, here the to-be-remembered words were unrelated to the sentences and were presented separately. That is, the participant would read a sentence aloud, and then an unrelated word would be presented, and so on until the end of a list, at which point the participant was required to recall the to-be-remembered words in correct serial order. This manipulation was introduced in an attempt to divorce the memoranda from the sentence context. If phonological similarity itself serves as a retrieval cue, facilitation should still be observed.

Another motivating factor for E2 was that the phonological similarity facilitation effect observed by Copeland and Radvansky (2001) was stronger than the facilitation effect observed in the present E1. An analysis revealed that our E1 sentences had slightly fewer words and syllables, on average, than did those used by Copeland and Radvansky. Nairne and Kelly (1999) demonstrated that increasing the delay between encoding and recall produces a shift from phonological similarity decrement to facilitation, so it is possible that the (cumulative) length and/or complexity of the sentences could impact the phonological similarity effect in reading span. We therefore administered two versions of reading span, one consisting of relatively short sentences and one consisting of relatively long sentences.

Method

Participants

Forty undergraduate students participated in exchange for partial course credit. Participants were assigned to

one of two reading span tasks: reading span A (shorter sentences similar to those used in E1 [$n = 20$]) or reading span B (longer sentences used by Copeland & Radvansky [$n = 20$]). The participants were all native English speakers.

Materials

The to-be-remembered words were the same as those used in E1.

Reading span A Sentences in this version of reading span were modified sentences from E1. The final word in each sentence was changed to another conceptually and grammatically acceptable word, if possible. Some sentences needed to be further altered if the final word was not easily substitutable. The to-be-remembered words were presented separately from the sentences and were the same words as those used in E1. The sentences and words were paired such that a to-be-remembered word was not paired with the sentence it came from in E1, again to ensure that there was no relationship between the sentences and the words.

Reading span B Sentences in this version of reading span were taken from Copeland and Radvansky (2001). Since their sentences were constructed to have some sentence-final words that were phonologically similar, the sentences were rearranged such that sentences with phonologically similar final words were not presented within the same list. As in reading span A, the to-be-remembered words were presented separately from the sentences and were the same words as those used in E1.

Procedure

The procedure was the same as that in E1, except that the experimenter clicked the mouse to present the to-be-remembered word immediately after the participant had finished reading the previous sentence. The to-be-remembered word remained on the screen for 1,000 ms, and then the next sentence or the recall screen was presented.

Scoring

The scoring procedures were the same as those in E1.

Results and discussion

Significance tests and effect sizes were conducted in the same manner as in E1. As a manipulation check, an

ANOVA was conducted to compare average reading time per sentence in the two reading span tasks. As was expected, participants were faster to read the shorter sentences in reading span A than the longer sentences in reading span B ($M_A = 3,851$ ms, $M_B = 4,504$ ms), $F(1, 38) = 49.62$, $p < .01$, $\eta^2 = .37$.

Serial recall A 2×2 mixed factorial ANOVA was conducted to examine the effect of phonological similarity on serial recall in the two reading span tasks. Phonological similarity facilitation was observed, replicating the results from E1. That is, the main effect of phonological similarity was significant, $F(1, 38) = 16.49$, $p < .01$, $\eta_p^2 = .30$, such that participants recalled more words in the similar condition than in the dissimilar condition (see Fig. 3a). The main effect of task was not significant, $F(1, 38) = 0.20$, $p = .66$, $\eta_p^2 = .01$, nor was the interaction, $F(1, 38) = .11$, $p = .75$, $\eta_p^2 = .00$.

In sum, participants demonstrated phonological similarity facilitation in both versions of reading span, suggesting that, as before, sentence context is not driving the phonological similarity facilitation effect. These results also indicate that our sentence length manipulation had no impact on the facilitation effect.

Item recall A 2×2 mixed factorial ANOVA was also conducted on item recall. Phonological similarity facilitation was again observed, replicating the results from E1. That is, the main effect of phonological similarity was significant, $F(1, 38) = 56.03$, $p < .00$, $\eta_p^2 = .60$, such that participants recalled more words in the similar condition than in the dissimilar condition (see Fig. 3b). The main effect of task was not significant, $F(1, 38) = 0.58$, $p = .45$, $\eta_p^2 = .02$, nor was the interaction, $F(1, 38) = 0.48$, $p = .49$, $\eta_p^2 = .01$.

Order accuracy A 2×2 mixed factorial ANOVA was also conducted on order accuracy. None of the effects in this analysis were significant (see Fig. 3c)[similarity, $F(1, 38) = 2.39$, $p = .13$, $\eta_p^2 = .06$; task, $F(1, 38) = 0.25$, $p = .62$, $\eta_p^2 = .01$; interaction, $F(1, 38) = 2.61$, $p = .12$, $\eta_p^2 = .06$].

To summarize the results of E2, phonological similarity facilitation was observed in both versions of reading span, in serial recall and in item recall. No effects were observed for order accuracy. All of these results replicate the results of E1.

Individual differences in facilitation in E1 and E2 Taken together, the results of the first two experiments demonstrated robust phonological similarity facilitation in four different versions of reading span. Here, we examined individual differences in the magnitude of the facilitation

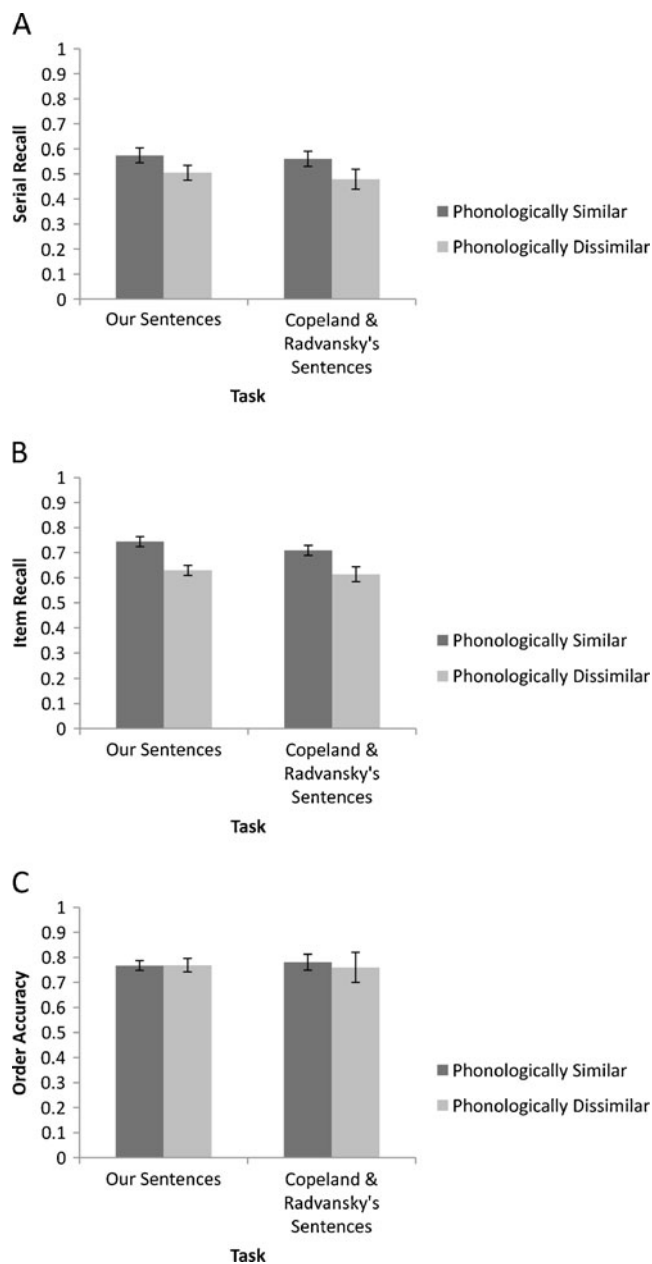


Fig. 3 Serial recall, item recall, and order accuracy in Experiment 2. Words within phonologically similar lists rhymed. Error bars represent the standard error of the mean

effect. Unsworth and Engle (2007) suggested that individuals with low WMC are not as efficient at performing a discriminative search of items in secondary memory and are more likely to include irrelevant items, such as items from previous lists. According to this view, low-WMC individuals should benefit more from a list retrieval cue that delimits the search set than should high-WMC individuals. Assuming that phonological similarity in E1 and E2 served as a list retrieval cue, the lower one's WMC, the better the participant should have performed in the phonologically

similar condition, relative to the dissimilar condition. Reading span tasks are typically administered with dissimilar items, and reading span is a reliable and valid measure of WMC, so each individual participant's WMC was calculated on the basis of his or her score from the dissimilar condition. Individual phonological similarity facilitation scores were calculated by subtracting recall in the phonologically dissimilar condition from recall in the phonologically similar condition. There was a significant correlation between WMC and the facilitation score, ($r = .44, p < .01$, suggesting that individuals with lesser WMC benefited more from phonological similarity (see Fig. 4).

Experiment 3

A third experiment was conducted in order to further test rhyming as a list retrieval cue. The results from the first two experiments supported the list context hypothesis, and in the first two experiments, phonologically similar lists consisted of words that rhyme. It is likely that rhyme served as a retrieval cue, which resulted in facilitation in reading span. In Experiment 3 (E3) phonological similarity was operationalized as feature overlap rather than rhyme. That is, words within similar lists shared phonological features but did not rhyme. The experiment was a 2×2 mixed factorial design in which participants performed either a word span or a reading span task and phonological similarity was manipulated within groups. The word span task was similar to word span in E1. The reading span task was similar to those in E2, in that the to-be-remembered words were unrelated to the sentences and were presented separately from the sentences.

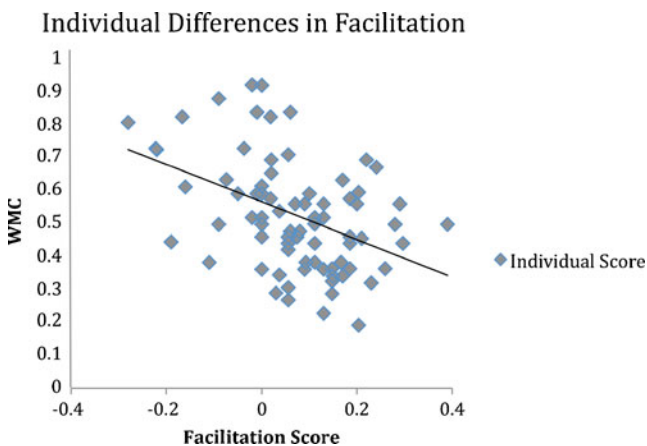


Fig. 4 Correlation between working memory capacity and phonological similarity facilitation, Experiments 1 and 2

Method

Participants

Forty-one undergraduate students participated in E3 in exchange for partial course credit. Participants were randomly assigned to either a simple span task or a complex span task. Participants assigned to the simple span task completed word span ($n = 20$). Participants assigned to the complex span task completed reading span ($n = 21$). The participants were all native English speakers.

Materials

A total of 108 single-syllable words were used as memoranda. Fifty-four of the words were arranged in three lists of 3, 4, 5, and 6 such that the words within each list were phonologically similar to one another but did not rhyme (e.g., *cap, man, cat*). The other 54 words were arranged in three lists of 3, 4, 5, and 6 such that the words within each set were phonologically dissimilar from one another (e.g., *zip, mop, jaw*). These words were taken primarily from Gupta, Lipinski, and Aktunc's (2005) list of canonically similar items and dissimilar items, with additional words taken from Oberauer's (2009) phonologically similar and phonologically dissimilar items and Baddeley's (1966) phonologically similar and dissimilar words. Sentences presented in the reading span task were the same sentences as those used in reading span A in E2.

Procedure

The procedure for word span was the same as that in E1, and the procedure for reading span was the same that in as E2.

Scoring

The same scoring procedures were used as those in the previous experiments.

Results and discussion

All statistical analyses were conducted in the same manner as in the previous experiments.

Serial recall A 2×2 mixed factorial ANOVA was conducted to examine the effect of phonological similarity on serial recall in the two tasks. The main effect of phonological similarity was significant, $F(1, 39) = 6.86, p = .01, \eta_p^2 = .15$, such that participants recalled more words in the dissimilar condition than in the similar

condition. The main effect of task was also significant, $F(1, 39) = 32.06, p < .01, \eta_p^2 = .45$, such that participants recalled more words in the word span task than in the reading span task. However, these main effects were qualified by a significant interaction, $F(1, 39) = 12.73, p < .01, \eta_p^2 = .25$ (see Fig. 5a). Simple effects analyses revealed the classic phonological similarity decrement in word span, $F(1, 19) = 24.78, p < .01, \eta^2 = .57$, but no effect of similarity in reading span, $F(1, 20) = 0.37, p = .55, \eta^2 = .02$.

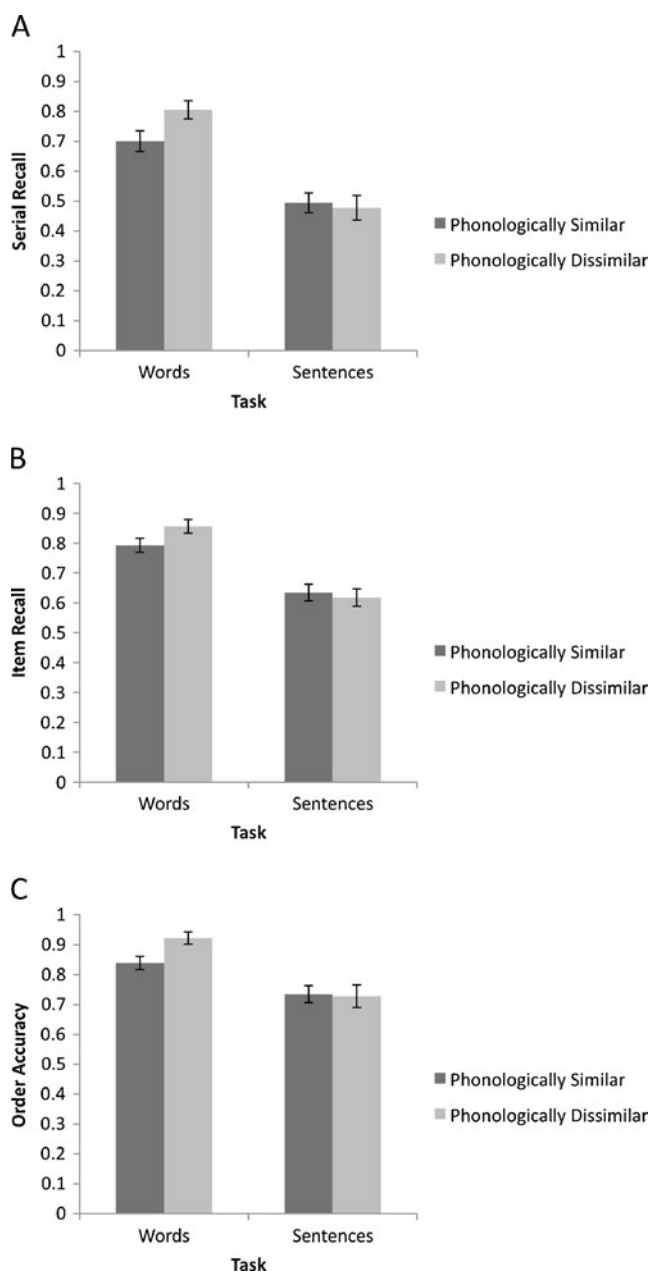


Fig. 5 Serial recall, item recall, and order accuracy in Experiment 3. Words within phonologically similar lists shared phonemic features but did not rhyme. Error bars represent the standard error of the mean

Phonological similarity facilitation was not observed in reading span when similarity was operationalized as feature overlap. This result supports our interpretation of the previous facilitation effects; that is, facilitation occurs because rhyming words provide a retrieval cue. It is also important to note that mean performance in the dissimilar condition of the various reading span tasks across experiments was remarkably consistent—E1 reading span, $M = .49$; E2 short reading span, $M = .48$; E2 long reading span, $M = .50$; E3 reading span, $M = .48$ —suggesting that phonological similarity facilitation is not a function of task difficulty.

Item recall A 2×2 mixed factorial ANOVA was also conducted on item recall. The main effect of phonological similarity was not significant, $F(1, 39) = 3.23, p = .08, \eta_p^2 = .08$, but the main effect of task was significant, $F(1, 39) = 32.57, p < .01, \eta_p^2 = .46$, such that participants recalled more words in the word span task than in the reading span task. As with serial recall, the interaction was significant, $F(1, 39) = 9.46, p < .01, \eta_p^2 = .20$ (see Fig. 5b). The simple effects were also consistent with the serial recall data. There was a phonological similarity decrement in word span, $F(1, 19) = 14.36, p < .01, \eta^2 = .43$, but no effect of similarity in reading span, $F(1, 20) = 0.71, p = .41, \eta^2 = .03$.

Order accuracy A 2×2 mixed factorial ANOVA was also conducted on order accuracy. The pattern of results was the same as that for item recall (see Fig. 5c). The main effect of phonological similarity was not significant, $F(1, 39) = 3.51, p = .07, \eta_p^2 = .08$, but the main effect of task was significant, $F(1, 39) = 18.87, p < .01, \eta_p^2 = .33$, such that order accuracy was higher in the word span task than in the reading span task. Again, the interaction was significant, $F(1, 39) = 4.89, p = .03, \eta_p^2 = .11$. Similar to E1, there was a phonological similarity decrement in word span, $F(1, 19) = 17.96, p < .01, \eta^2 = .49$, but no effect of similarity in reading span, $F(1, 20) = 0.04, p = .85, \eta^2 = .00$.

To summarize the results of E3, the classic phonological similarity decrement was observed in word span, across all measures of performance. In contrast, nonrhyming phonological similarity had no effect on reading span.

General discussion

The present experiments demonstrate that simple and complex span tasks are sensitive to phonological similarity, but in different ways. The classic phonological similarity decrement was observed in simple span, regardless of how similarity was operationalized. In contrast, phonological

similarity facilitation was observed in complex span, but only when similar lists consisted of rhyming words, not when similarity was operationalized as feature overlap. The experiments do not support the hypothesis that phonological similarity facilitation in reading span occurs because of sentence context (Copeland & Radvansky, 2001). In E1, we enhanced context within lists in one of the conditions, and in E2, we removed the memoranda from the sentence context. Neither increasing nor decreasing context altered phonological similarity facilitation.

The present results are consistent with psychometric and neuroimaging research that has similarly demonstrated empirical dissociations between simple and complex span tasks. While the data suggest an important and fundamental difference between simple and complex span tasks, it would be incorrect to interpret these dissociations to mean that simple and complex span tasks tap completely different cognitive and neural mechanisms. Clearly, the two classes of tasks have much in common. Moreover, psychometric studies tend to show moderate to strong correlations between simple and complex span tasks. As well, neuroimaging experiments show that certain brain regions, such as the lateral prefrontal cortex and posterior parietal cortex, are recruited in both simple and complex span tasks but are *more* active in complex than in simple span.

The data therefore suggest that simple and complex span tasks tap the same set of cognitive mechanisms, but do so to different degrees. This interpretation is consistent with Unsworth and Engle (2007), who have argued that in complex span tasks, the memoranda are displaced from primary memory into secondary memory due to interference from the secondary task, which is being performed during encoding and maintenance of the to-be-remembered items (e.g., reading sentences aloud in reading span). This displacement and interference increases difficulty, lowering overall recall, and engenders the need for a discriminative search of secondary memory at the end of a list when the recall prompt is presented. According to this view, the classic phonological similarity *decrement* in simple span occurs because these tasks emphasize active maintenance in primary memory. Concurrently maintaining multiple items in primary memory that are phonologically similar results in increased confusability and lower recall, whereas phonologically dissimilar items in the same state do not suffer as much interitem interference. Phonological similarity *facilitation* occurs in complex span tasks because these tasks emphasize discriminative search of secondary memory. Quickly displacing phonologically similar memoranda into secondary memory allows the search set to be delimited to relevant items, thereby enhancing recall. Phonologically dissimilar items do not reap the benefit from a unifying list retrieval cue.

The present results are also consistent with Nairne's feature model of memory and, therefore, do not necessitate the notion of "active maintenance of items" (see Nairne, 1990, 2002). According to this perspective, all memory is cue driven, and what differs between immediate and delayed retention is simply the constellation of cues that drive retrieval. Given the differences between simple and complex span tasks, in terms of time from encoding to recall and time between encoding each subsequent item, it is reasonable to conclude that in simple span tasks, participants rely more upon phonological cues, whereas in complex span, participants rely more upon temporal, list, and/or semantic cues.

The question remains as to why Copeland and Radvansky (2001) did not obtain phonological similarity facilitation in the operation span task, which is a complex span task, similar in structure to reading span. They observed no effect of phonological similarity—that is, neither a significant decrement nor facilitation. We see at least three possible explanations of this result. First, it is possible that reading sentences aloud, as required in reading span, disrupts rehearsal and/or interferes more with phonological representations than does solving math problems, as required in operation span. A second possibility is that the act of reading sentences "primes" participants to activate semantic features of the to-be-remembered stimuli, resulting in more semantic representations in reading span, relative to operation span. Third, and perhaps least interesting, is that their null result is simply a Type II error. One aspect of their data suggests that this is possible. Oddly, recall in operation span was equivalent to recall in word span. This is almost never observed in experiments that include both simple and complex span tasks with the same to-be-remembered stimuli. Instead, simple span performance is typically greater than complex span performance. This aspect of their data makes the null result difficult to interpret and suggests that more experiments are needed to explore the effect of phonological similarity in operation span.

In conclusion, three experiments were conducted to investigate the effect of phonological similarity on simple and complex memory span tasks. The classic phonological similarity effect was observed in simple span, such that phonologically similar lists were recalled worse than phonologically dissimilar lists. In contrast, phonological similarity facilitation was observed in reading span when similar lists consisted of words that rhyme, but not when similar lists consisted of words that shared phonological features but did not rhyme. We interpret these results to suggest that rhyming created a list retrieval cue and cue-dependent memory retrieval mechanisms are more important for performance of complex span than of simple span.

Appendix

Table 1 Total error frequency by experimental condition and type of error

	S ^a	B ^b	PLI ^c	SI ^d	PI ^e	R ^f	O ^g
Phonologically Similar Condition							
Rhyming							
Experiment 1: Words	240	137	0	N/A	19	8	10
Experiment 1: Sentences	234	238	0	1	13	2	3
Experiment 1: Stories	217	195	0	0	10	2	6
Experiment 2: Our sentences	179	249	0	1	20	1	2
Experiment 2: 2001 sentences ^h	155	293	0	0	19	1	1
Nonrhyming							
Experiment 3: Words	120	150	1	N/A	61	4	5
Experiment 3: Sentences	157	332	8	7	58	0	8
Phonologically Dissimilar Condition							
Experiment 1: Words	93	70	15	N/A	17	3	12
Experiment 1: Sentences	162	335	9	14	5	0	6
Experiment 1: Stories	151	282	5	31	0	0	5
Experiment 2: Our sentences	337	20	4	9	0	21	
Experiment 2: 2001 sentences ^h	148	383	14	8	6	1	6
Experiment 3: Words	64	82	26	N/A	38	0	1
Experiment 3: Sentences	161	343	21	6	58	0	5

^a Sequence errors: Instances in which a participant correctly recalled an item in the incorrect position

^b Blank: Instances in which a participant left a blank space in place of a word

^c Previous list intrusion: Instances in which a participant erroneously recalled an item from a previous list

^d Sentence intrusion: Instances in which a participant erroneously recalled a word from a read sentence

^e Phonological interference: Instances in which a participant erroneously recalled a word that rhymed with a target word

^f Repeats: Instances in which a participant typed a single to-be-remembered item a second time in the same list

^g Other: Miscellaneous errors

^h Sentences created by Copeland and Radvansky (2001)

References

- Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 2, pp. 89–195). New York: Academic Press.
- Baddeley, A. D. (1966). Short-term memory for word sequences as a function of acoustic, semantic, and formal similarity. *The Quarterly Journal of Experimental Psychology*, *18*, 362–365.
- Botvinick, M., & Plaut, D. C. (2006). Short-term memory for serial order: A recurrent neural network model. *Psychological Review*, *113*, 201–233.
- Brown, J. (1958). Some tests of the decay theory of immediate memory. *The Quarterly Journal of Experimental Psychology*, *10*, 12–21.
- Bunge, S. A., Klingberg, T., Jacobsen, R. B., & Gabrieli, J. D. E. (2000). A resource model of the neural basis of executive working memory. *Proceedings of the National Academy of Sciences*, *97*, 3573–3578.
- Chein, J. M., Moore, A. B., & Conway, A. R. A. (2011). Domain-general mechanisms of complex working memory span. *Neuroimage*, *54*, 550–559.
- Coleman, M., & Liao, T. L. (1975). A computer readability formula designed for machine scoring. *The Journal of Applied Psychology*, *60*, 283–284.
- Coltheart, M. (1981). The MRC Psycholinguistic Database. *The Quarterly Journal of Experimental Psychology*, *33A*, 497–505.
- Conway, A. R. A., Cowan, N., Bunting, M. F., Theriault, D. J., & Minkoff, S. R. B. (2002). A latent variable analysis of working memory capacity, short-term memory capacity, processing speed, and general fluid intelligence. *Intelligence*, *30*, 163–183.
- Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, *12*, 769–786.
- Conway, A. R. A., Kane, M. J., & Engle, R. W. (2003). Working memory capacity and its relation to general intelligence. *Trends in Cognitive Sciences*, *7*, 547–552.
- Copeland, D. E., & Radvansky, G. E. (2001). Phonological similarity in working memory. *Memory & Cognition*, *29*, 774–776.

- Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *The Behavioral and Brain Sciences*, 24, 97–185.
- Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, 19, 450–466.
- Daneman, M., & Merikle, P. M. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin & Review*, 3, 422–433.
- Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. A. (1999). Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology: General*, 128, 309–331.
- Flesch, R. (1948). A new readability yardstick. *The Journal of Applied Psychology*, 32, 221–233.
- Gilhooly, K. J., & Logie, R. H. (1980). Age of acquisition, imagery, concreteness, familiarity and ambiguity measures for 1944 words. *Behavior Research Methods & Instrumentation*, 12, 395–427.
- Gunning, R. (1952). *The technique of clear writing*. New York: McGraw-Hill.
- Gupta, P., Lipinski, J., & Aktunc, E. (2005). Reexamining the phonological similarity effect in immediate serial recall: The roles of type of similarity, category cuing, and item recall. *Memory & Cognition*, 33, 1001–1016.
- Jonides, J., Lewis, R. L., Nee, D. E., Lustig, C. A., Berman, M. G., & Moore, K. S. (2008). The mind and brain of short-term memory. *Annual Reviews of Psychology*, 59, 193–224.
- Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The generality of working memory capacity: A latent variable approach to verbal and visuo-spatial memory span and reasoning. *Journal of Experimental Psychology: General*, 133, 189–217.
- Kincaid, J. P., Fishburne, R. P., Jr., Rogers, R. L., & Chissom, B. S. (1975). *Derivation of new readability formulas (Automated Readability Index, Fog Count and Flesch Reading Ease Formula) for Navy enlisted personnel (Research Branch Report 8–75)*. Millington, TN: Naval Technical Training, U. S. Naval Air Station, Memphis, TN.
- Kondo, H., Morishita, M., Osaka, N., Osaka, M., Fukuyama, H., & Shibasaki, H. (2004). Functional roles of the cingulo-frontal network in performance on working memory. *Neuroimage*, 21, 2–14.
- Kučera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University.
- Lewandowsky, S., & Farrell, S. (2008). Phonological similarity in serial recall: Constraints on theories of memory. *Journal of Memory and Cognition*, 58, 429–448.
- McLaughlin, G. H. (1969). SMOG Grading—A new readability formula. *Journal of Reading*, 12, 639–646.
- Nairne, J. S. (1990). A feature model of immediate memory. *Memory & Cognition*, 18, 251–269.
- Nairne, J. S. (2002). Remembering over the short-term: The case against the standard model. *Annual Review of Psychology*, 53, 53–81.
- Nairne, J. S., & Kelley, M. R. (1999). Reversing the phonological similarity effect. *Memory & Cognition*, 27, 45–53.
- Neale, K., & Tehan, G. (2007). Age and redintegration in immediate memory and their relationship to task difficulty. *Memory & Cognition*, 35, 1940–1953.
- Oberauer, K. (2009). Interference between storage and processing in working memory: Feature overwriting, not similarity-based competition. *Memory & Cognition*, 37, 346–357.
- Osaka, M., Osaka, N., Kondo, H., Morishita, M., Fukuyama, H., Aso, T., et al. (2003). The neural basis of individual differences in working memory capacity: An fMRI study. *Neuroimage*, 181, 789–797.
- Osaka, N., Osaka, M., Kondo, H., Morishita, M., Fukuyama, H., & Shibasaki, H. (2004). The neural basis of executive function in working memory: An fMRI study based on individual differences. *Neuroimage*, 21, 623–631.
- Page, M., & Norris, D. (1998). The primacy model: A new model of immediate serial recall. *Psychological Review*, 105, 761–781.
- Pavio, A. (2011). *Familiarity, concreteness, and imageability norms for words*. Unpublished manuscript.
- Peterson, L. R., & Peterson, M. J. (1959). Short-term retention of individual verbal items. *Journal of Experimental Psychology*, 58, 193–198.
- Smith, E. E., Geva, A., Jonides, J., Miller, A., Reuter-Lorenz, P., & Koeppel, R. A. (2001). The neural basis of task-switching in working memory: Effects of performance and aging. *Proceedings of the National Academy of Sciences*, 98, 2095–2100.
- Tehan, G., Hendry, L., & Kocinski, D. (2001). Word length and phonological similarity effects in simple, complex and delayed serial recall tasks: Implications for working memory. *Memory*, 9, 333–348.
- Toglia, M. P., & Battig, W. R. (1978). *Handbook of semantic word norms*. Hillsdale, NJ: Erlbaum.
- Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language*, 28, 127–154.
- Unsworth, N., & Engle, R. W. (2007). The nature of individual differences in working memory capacity: Active maintenance in primary memory and controlled search from secondary memory. *Psychological Review*, 114, 104–132.

Author Note

We would like to thank Sarah Hoffman and Julianna Stevens for assistance with data collection. We would also like to thank David Copeland for providing stimulus materials for Experiment 2, which were previously used by Copeland and Radvansy (2001).