The roles of semantic similarity and proactive interference in the word length effect

WINSTON D. GOH and CHANG KHIANG GOH
National University of Singapore, Singapore

We examined the influence of semantic similarity and proactive interference (PI) on the word length effect (WLE) in immediate serial recall. Word length was manipulated by comparing memory for monosyllabic versus multisyllabic words. PI effects were evaluated by manipulating semantic similarity in the to-be-remembered lists and examining its impact on the WLE's magnitude across eight-trial blocks. Words were sampled from a single semantic category across the entire block, from a single category within the list, or from different categories. Robust WLEs were observed in single-category blocks and when words were from different categories. However, when all the within-list words were from the same semantic category, the WLE was sharply attenuated. Except for the within-list semantic similarity condition, there was a buildup in PI levels in the form of protrusion errors across trials. However, the magnitude of the WLE did not increase with the PI buildup, suggesting that it was not affected by PI across trials.

The word length effect (WLE) refers to better recall for short versus long words in short-term memory (STM; Baddeley, Thomson, & Buchanan, 1975). The standard account for this phenomenon has been framed in terms of decay-based theories of immediate memory performance (Baddeley & Hitch, 1974), in which the short-word advantage is attributed to the ability to rehearse short words more often than long words per unit time. Thus, rehearsal mechanisms are implicated in the trace maintenance processes employed to counter decay.

This decay-based explanation has been challenged by research showing inconsistent effects of spoken duration on recall (Lovatt, Avons, & Masterson, 2000, 2002; Neath, Bireta, & Suprenant, 2003) and by demonstrations that phonological complexity (Caplan, Rochon, & Waters, 1992; Service, 1998), rather than articulatory duration, influences the WLE. However, Mueller, Seymour, Kieras, and Meyer (2003) have suggested that with more precise measures of articulatory durations, it can be shown that time-based explanations may be viable. Other researchers have suggested that distinctiveness and complexity factors also need to be taken into account (Cowan, Baddeley, Elliott, & Norris, 2003; Hulme, Surprenant, Bireta, Stuart, & Neath, 2004).

Neath and Nairne (1995) have proposed an alternative, interference-based account of the WLE. In this account, words are represented by a number of segments, with long words assumed to contain more segments than do short words. During recall, these segments need to be reconstructed from degraded traces in STM, which are

then compared with an appropriate search set in long-term memory (LTM) for retrieval. There is a fixed probability of an assembly error for each segment. Because long words have more segments than do short words, the probability of reconstruction errors is greater for the former, which then reduces the efficacy of the trace as a retrieval cue by which to discriminate items in the memory search set. Hence, the probability of successfully recalling a long word is worse than that of recalling a short word.

Nairne, Neath, and Serra (1997) also suggested that proactive interference (PI) may influence the WLE. This serial recall study of five-word lists, with word length manipulated by spoken duration, demonstrated that the WLE emerged only after the first four-trial block. Nairne et al. argued that during the initial trials, when there were fewer items in STM, participants were able to retrieve trial information from LTM. However, retrieval from LTM becomes difficult after several trials, since the current trial cannot be easily discriminated from previous trials, due to increasing PI. The absence of word length differences in the initial trials is damaging to decay explanations, because decay per se should not depend on trial sequence. Word length rehearsal differences should be evident across all trials, not just the later ones.

Recently, Tehan and Turcotte (2002) examined the role of PI in the WLE in more detail. In Experiment 1, they replicated Nairne et al.'s (1997) finding of no WLE, using a spoken duration manipulation. However, the WLE was evident when monosyllabic words were compared with multisyllabic words. Tehan and Turcotte (2002) suggested that syllable length, rather than spoken duration, was a better manipulation of word length, and this might explain why a reliable WLE was observed with a syllable length manipulation (see also Tolan & Tehan, 2005). In Experiment 2, PI influences on the WLE were examined using a modification of Tehan and Humphreys's (1996) cued recall task. In this serial recall version, the critical trials in-

This work was supported by Research Grants C-107-000-222-091 and R-581-000-048-112 to W.D.G. We thank Elisabeth Service, Gerald Tehan, and an anonymous reviewer for constructive criticisms. The word lists are available from the authors. Address correspondence to W. D. Goh, Department of Psychology, National University of Singapore, Singapore 117570 (e-mail: psygohw@nus.edu.sg).

volved discriminating items in a to-be-remembered block from those in a prior distractor block. PI was manipulated by varying the nature of the items in the distractor block: consonants (no interference) versus words (interference). A WLE was found, but there was no evidence of interference effects. On the basis of the overall findings, Tehan and Turcotte concluded that there was no evidence of PI influences on the WLE.

Nairne et al. (1997) observed a reliable WLE from the second four-trial block onward, whereas Tehan and Turcotte (2002) used a single four-trial block and demonstrated that the WLE was evident when a syllable length manipulation was used. It is possible that the WLE's *magnitude* would have increased had more trials been included. This would be consistent with Nairne et al.'s argument that PI effects build up over trials and contribute to the WLE. In other words, PI influences can be implicated if the WLE's magnitude increases proportionately with trial sequence.

We suggest that a stronger test of PI can be made with semantic manipulations. Semantic attributes, such as taxonomic categories, have the highest efficacy in obtaining a release-from-PI effect (Wickens, 1970). Consequently, semantic dimensions should be effective in differentiating PI levels. Poirier and Saint-Aubin (1995) showed that serial recall of lists comprising words from the same semantic category (homogeneous list) was better than that of lists with words from different categories (heterogeneous list). Semantic organization provided an extra retrieval cue that restricted the memory search set size to categorically appropriate words and should minimize PI effects. Poirier and Saint-Aubin's homogeneous list manipulated within-list homogeneity; that is, words in one list were from one category (e.g., *animals*), whereas words in the next switched to a different category (e.g., *fruits*), and so on. Across-list PI should be minimal in such circumstances. We suggest that maximal PI effects may be observed in a condition in which all the lists in a *block* were from a single category (homogeneous block), which will be analogous to the PI buildup trials in the release-from-PI paradigm.

The goal of the present study was to extend Nairne et al.'s (1997) and Tehan and Turcotte's (2002) research by determining the extent to which PI influences on the WLE could be demonstrated with the strong syllabic manipulation of word length. We included eight trials per block because Nairne et al. demonstrated robust WLEs at Trials 5–8. We also adapted Poirier and Saint-Aubin's (1995) manipulation of semantic similarity in three conditions: homogeneous block, in which all the words were from one category; homogeneous list, in which the within-list words were from the same category; and heterogeneous list, in which the words were from different categories.

EXPERIMENT 1

Method

Participants. Fifty introductory psychology students participated for course credit.

Design and Materials. A 2 (length) $\times$ 3 (similarity) $\times$ 8 (trial) completely within-subjects design was used. Although Tehan and

Turcotte (2002) reported no WLE differences between closed and open sets of words, we used an open set, following previous research in which PI was manipulated with taxonomic categories (Wickens, 1970). La Pointe and Engle (1990) also reported more robust WLEs with open sets. Four hundred sixteen words forming 18 categories were selected, with most being taken from Battig and Montague (1969). Half were short words (195 monosyllabic and 13 disyllabic) and half were long (192 trisyllabic and 16 with four syllables).

For the homogeneous blocks, it was impossible to select 40 (8 trials $\times$ 5 words) short and 40 long words from one category in Battig and Montague (1969) while balancing word frequency. Thus, closely related categories (e.g., *fruits* and *vegetables*) were combined to create the homogeneous blocks. Some of the 16 categories used for the homogeneous lists and heterogeneous lists were also derived from combinations of related categories (e.g., *alcoholic* and *nonalcoholic beverages*). These 16 categories contained 8 words each.

Two word list versions were created for counterbalancing. In one version, the homogeneous block's short words were from the *animals* category, whereas long words were *plants*. One half of the other 16 categories supplied short words for the homogeneous lists and long words for the heterogeneous lists. The remaining 8 categories supplied long words for the homogeneous lists and short words for the heterogeneous lists. In the other version, the sampling was reversed; that is, the homogeneous block's short words were *plants* and the long words were *animals*, and so on. The average word frequencies of the short and long words in the versions were balanced.

Procedure. A balanced Latin square was used to counterbalance the six-block presentation sequence making up the length $\times$ similarity conditions. Each block comprised eight trials of five-word lists. The participants were randomly assigned to one of the six sequences derived from the Latin square and to each word list version.

The participants were tested in groups, using individual PCs. A trial sequence consisted of a "ready" prompt centered on the monitor, a 500-msec blank screen, five words visually presented one at a time at a rate of one word per second, and a "recall" prompt. The participants performed strict serial recall on prepared answer sheets. The next trial was self-initiated with a keypress. At the end of each block, the participants performed a 90-sec mathematical task to minimize carryover effects.

Within each trial, the 5-word presentation sequences were randomized, and sampling was performed without replacement. In the homogeneous blocks, 5 words were randomly sampled from the relevant pool of 40 short or long words. In the homogeneous lists, the category for each trial was randomly sampled from eight categories, and 5 words were randomly sampled from 8 words within that category. In the heterogeneous lists, five categories were randomly sampled from eight categories, and 1 word was randomly sampled from each selected category.

The participants were given two practice trials comprising words unrelated to the experiment at the beginning of the study.

Results and Discussion

Performance was scored using correct-in-position (serial recall) and pure item recall criteria. Separate three-way ANOVAs and the appropriate simple effects were run for each scoring method, using a .05 alpha level. The influence of PI and semantic similarity on the WLE should be manifested in interactions involving both or either of the similarity and trial factors with length, so we will focus on these interactions. Figure 1 depicts recall performance across conditions.

Serial recall. The main effect of trial was significant [$F(7,343) = 2.63$, $MS_e = 0.05$], with a linear trend indicating a general decline in performance across trials

[$F(1,49) = 11.71$, $MS_e = 0.06$]. Main effects of length [$F(1,49) = 73.97$, $MS_e = 0.09$] and similarity [$F(2,98) = 17.79$, $MS_e = 0.06$] were observed, but their interpretation was superseded by a significant length $\times$ similarity interaction [$F(2,98) = 17.22$, $MS_e = 0.05$].

Table 1 summarizes the nature of this interaction. The short-word advantage was reliable in all the similarity conditions [$ts(49) > 2.2$], but the WLE was clearly attenuated with the homogeneous lists; more long words were recalled here than in the other two conditions [$ts(49) > 5.9$], with no reliable difference between homogeneous blocks and heterogeneous lists ($t < 1$). For short words, homogeneous blocks and homogeneous lists were better recalled than heterogeneous lists [$ts(49) > 2.4$], but there was, unexpectedly, no difference between homogeneous blocks and homogeneous lists [$t(49) = 1.61$].¹

No other interactions were significant ($F_s < 1$). The lack of any interactions involving trial and length suggests that the WLE is not dependent on PI buildup.

Item recall. Only two results deviated from the serial recall findings. First, the homogeneous list WLE, previously attenuated but still reliable, was eliminated [$t(49) = 1.06$]. Second, the trial $\times$ similarity interaction was marginally significant [$F(14,686) = 1.52$, $MS_e = 0.03$, $p < .10$], which is attributable to linearly declining recall across trials with homogeneous blocks and heterogeneous lists [$F_s(1,49) > 9.2$, $MS_e < 0.03$], where PI buildup was expected, but not with homogeneous lists ($F < 1$), where minimal PI was expected.

Errors. Four error types are summarized in Table 2: protrusions (within-block recall from prior lists, a direct index of PI), transpositions, omissions, and item errors. Separate length $\times$ similarity ANOVAs were conducted for each type of error, followed by appropriate pairwise tests.

For protrusions, only a main effect of similarity was found [$F(2,98) = 20.20$, $MS_e = 0.000$]. Homogeneous blocks had more protrusions than did heterogeneous lists [$t(49) = 2.20$], indicating greater PI in the former, as was

expected. Unsurprisingly, there were practically no protrusions with homogeneous lists.² This is consistent with the marginal trial $\times$ similarity interaction observed with item recall, which was evidence for differential PI levels among the semantic conditions. Importantly, the lack of an interaction with length suggests no differential protrusions between short and long words, indicating once again that the WLE does not depend on PI.³

Trade-offs between transposition and item errors for short and long words were observed. No interactions were reliable, but main effects of length were observed [$F_s(1,49) > 15.0$, $MS_e < 0.005$], with more long-word than short-word transpositions across all conditions and the reverse trend for item errors. We will defer commentary on these effects to the General Discussion section. There was also a main effect of similarity for item errors [$F(2,98) = 8.95$, $MS_e = 0.002$], with the fewest errors for homogeneous blocks, but no main effect of similarity for transpositions ($F < 1.13$), which is consistent with research showing that semantic similarity does not affect order information (Nairne & Kelley, 2004; Saint-Aubin & Poirier, 1999).

Omissions mirrored the recall results, with the length $\times$ similarity interaction [$F(2,98) = 17.66$, $MS_e = 0.004$] attributable to more long- than short-word omissions in all the conditions [$ts(49) > 4.7$], but with the fewest long- and short-word omissions in homogeneous lists [$ts(49) > 6.8$] and heterogeneous lists [$ts(49) > 3.5$], respectively. No other differences were significant ($ts < 1$).

The overall results support Tehan and Turcotte's (2002) observation that WLEs, when elicited by syllable length manipulations, are generally robust across all trials. No reliable interactions involving trials and word length were found, suggesting that the WLE's magnitude is not differentially affected by a PI buildup when the number of trials increases, as was found by Nairne et al. (1997). Within-list semantic similarity attenuated the WLE, suggesting that the additional categorical retrieval cue reduced the propensity of PI from prior trials by limiting the memory search set.⁴

EXPERIMENT 2

Before it is concluded that PI does not play a part in the WLE, the possibility should be considered that our within-subjects design was not an adequate test of PI buildup effects because of potential PI carryover effects across conditions and from practice trials. In Nairne et al. (1997), word length was run between subjects, with no practice trials. We observed little evidence of protrusions from practice trials or across blocks, but it is possible that such carryover effects could be manifested in other ways, such as omissions. Hence, we sought to determine whether the results would hold when length and similarity were run between subjects.

Method

Participants. Three hundred seventy-eight introductory psychology students participated for course credit.

Table 1
Recall Probability Collapsed Across Trials

Word Length	Semantic Similarity					
	Homogeneous Block		Homogeneous List		Heterogeneous List	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Experiment 1						
Serial recall						
Short	.72	.13	.69	.13	.65	.14
Long	.55	.14	.65	.12	.55	.12
Item recall						
Short	.75	.12	.73	.11	.69	.12
Long	.61	.12	.72	.10	.61	.10
Experiment 2						
Serial recall						
Short	.70	.13	.66	.15	.63	.14
Long	.54	.11	.60	.13	.51	.11
Item recall						
Short	.74	.12	.72	.12	.67	.13
Long	.61	.10	.69	.12	.59	.09

Table 2
Error Rates Across Semantic Similarity and Word Length

		Semantic Similarity					
Error Type	Word Length	Homogeneous Block		Homogeneous List		Heterogeneous List	
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Experiment 1							
Protrusions	Short	.02	.02	.00	.01	.01	.01
	Long	.02	.02	.00	.00	.01	.02
Transpositions	Short	.03	.04	.04	.05	.04	.04
	Long	.07	.06	.07	.07	.06	.05
Item	Short	.06	.06	.09	.07	.08	.06
	Long	.04	.04	.05	.06	.04	.04
Omissions	Short	.18	.10	.17	.10	.22	.13
	Long	.34	.12	.23	.10	.34	.10
Experiment 2							
Protrusions	Short	.02	.02	.00	.01	.01	.02
	Long	.03	.04	.00	.00	.01	.02
Transpositions	Short	.05	.05	.06	.05	.04	.04
	Long	.07	.06	.10	.06	.08	.05
Item	Short	.07	.05	.08	.06	.10	.08
	Long	.03	.03	.02	.03	.03	.03
Omissions	Short	.17	.11	.20	.11	.22	.10
	Long	.33	.11	.29	.11	.37	.10

Design and Procedure. Details were the same as those in Experiment 1, except that practice trials were omitted. The participants were randomly assigned to one between-subjects condition and an appropriate word list.

Results and Discussion

The data of 4 participants were discarded due to computer failure. Figure 1 illustrates the recall patterns, which were analyzed with the methods used previously. The patterns were essentially identical to those in Experiment 1, so we will report only the statistical values for the critical findings and any deviations.

Serial/item recall. All the main effects were observed but were superseded by two-way interactions. As is depicted in Table 1, across both scoring methods, the nature of the length $\times$ similarity interactions [$F_s(1,368) > 4.3$, $MS_e < 0.14$] was similar; the WLE was robust in all the conditions except homogeneous lists, in which it was attenuated with serial recall and eliminated with item recall. One deviation was that in Experiment 1, short-word recall was better with homogeneous lists than with heterogeneous lists with both scoring methods, but in Experiment 2, this difference was reliable only with item recall [$t(122) = 2.15$], and not with serial recall [$t(122) = 1.11$].

The trial $\times$ similarity interaction was marginally significant with serial recall [$F(14,2576) = 1.60$, $MS_e = 0.05$, $p < .08$] and was reliable with item recall [$F(14,2576) = 1.95$, $MS_e = 0.04$]. This was again attributable to linear trends for trials with homogeneous blocks and heterogeneous lists, but not for those with homogeneous lists. As before, no other interactions were reliable, including interactions between trial and length, which would be evidence that the WLE is dependent on PI buildup.

Errors. Table 2 summarizes the error patterns, which were similar to those in Experiment 1. The main effect of

similarity for protrusions [$F(2,368) = 22.72$, $MS_e = 0$] was due to the expected higher incidence of homogeneous block protrusions than of heterogeneous list protrusions [$t(247) = 3.49$]. The trade-off between transposition and item errors was also consistent with the results in Experiment 1: more transpositions for long words, but more item errors for short words [$F_s(1,368) > 38.0$, $MS_e < 0.004$].

There were several inconsistencies in the error results across experiments,⁵ but the overall trend and major findings were similar. The WLE was sharply attenuated by within-list semantic similarity. There was evidence of PI with homogeneous blocks and heterogeneous lists, but there was no corresponding evidence that the WLE's magnitude was affected by PI across trials.

GENERAL DISCUSSION

Does PI play a role in the WLE? In terms of the building up of PI over the course of multiple trials, there was no evidence to indicate that short- and long-word recall differences were magnified proportionally to increasing PI levels. There were also no differential short- and long-word protrusion levels.

In terms of preventing cross-list intrusions with additional retrieval cues such as category membership, it could be argued that under conditions in which PI was minimized, the normal WLE could be overcome by other factors. Accounts of PI and release from PI effects revolve around the continued efficacy of retrieval cues (Wickens, 1970). The idea is that when words continue to be sampled from the same category, the category retrieval cue's effectiveness diminishes as it becomes increasingly difficult to discriminate current trial items from previous ones. This is essentially what happened with homogeneous blocks and, to a smaller extent, with

Figure 1. Recall probability (+ SEs) across trials as a function of word length and semantic similarity in Experiment 1 (left panels) and Experiment 2 (right panels).

heterogeneous lists. In both cases, there was direct evidence of PI protrusions. When the category changed across homogeneous lists, there was essentially a release from PI at every trial, and virtually no protrusions occurred. The unique category supplied an extra retrieval cue that effectively restricted the size of the memory search set at the point of recall. Such cues were not helpful in the other conditions, because they matched items in prior trials. If the WLE occurs because the probability of reconstruction failure for long words is greater, due to longer segments (Neath & Nairne, 1995), providing cues that increase trace discriminability, relative to other LTM candidates, by limiting the search set would boost the probability of successful retrieval.

Our results also suggest a greater semantic similarity advantage for long words. The short-word advantage with homogeneous lists, relative to the other conditions, across both serial and item scoring was not robust. One possibility is that following Schweickert's (1993) two-process mechanism for recall, a direct read-out from STM is more prevalent for short words, thus obscuring similarity effects, whereas for long words, recall is more likely to involve reconstruction from degraded traces. Consequently, category cues are more likely to boost memory for long than for short words.

The error analyses indicate a consistent trade-off between short and long words for item and transposition errors. There were more extralist errors for short words and more transpositions for long words. Hendry and Tehan (2005) recently suggested that the WLE could be explained by an item order approach. In their account, within the typical presentation rates in STM studies, long words take longer to process and, therefore, receive greater item processing but have less time for order processing, whereas short words have more time for order processing. This is compatible with our result showing more long-word transpositions: The participants remembered the items but forgot their order. For short words, when item recall is erroneous, the likelihood of extralist intrusions is higher during trace reconstruction because it is easier to match other words that share phonological or other features with the degraded trace. Many item errors tended to be semantic and/or phonological substitutions—for example, *oak* for *oat*. This is consistent with previous findings regarding neighborhood errors in STM (Goh & Pisoni, 2003; Roodenrys, Hulme, Lethbridge, Hinton, & Nimmo, 2002). It is harder to make such substitutions for longer words, given that the number of matching competitors should be inversely proportional to word length. However, no error data were reported in Hendry and Tehan, so a direct comparison cannot be made with their word length error differences. It should be noted that the majority of errors in the present study were omissions, whose pattern mirrored the recall data. This basically suggests that most of the effects were due to differences in the extent of permanently degraded traces.

The present study does not support Nairne et al.'s (1997) results showing PI influences on the WLE. It may

be that the earlier study's closed set word pool generated PI at a faster rate and with a greater magnitude, since it is likely to be harder to discriminate item and order information if the same words are repeated across multiple trials. Although there was evidence of PI buildup in the present experiments, the PI increase could be at a much lower level with open set words.

In summary, our results extend and support Tehan and Turcotte's (2002) finding of robust WLEs, using a syllabic manipulation of word length. However, when extra retrieval cues, such as semantic similarity, are available at the point of recall, the WLE is sharply attenuated. These results are consistent with and add to the growing body of evidence suggesting that a cue-based approach to STM (Nairne, 2002) is a viable and promising approach to consider for explaining the underlying mechanisms responsible for the WLE and other STM effects.

REFERENCES

- BADDELEY, A. D., & HITCH, G. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 8, pp. 47-89). New York: Academic Press.
- BADDELEY, A. D., THOMSON, N., & BUCHANAN, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning & Verbal Behavior*, 14, 575-589.
- BATTIG, W. F., & MONTAGUE, W. E. (1969). Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology Monographs*, 80, 1-46.
- CAPLAN, D., ROCHON, E., & WATERS, G. S. (1992). Articulatory and phonological determinants of word length effects in span tasks. *Quarterly Journal of Experimental Psychology*, 45A, 177-192.
- 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, 74-79.
- GOH, W. D., & PISONI, D. B. (2003). Effects of lexical competition on immediate memory span for spoken words. *Quarterly Journal of Experimental Psychology*, 56A, 929-954.
- HENDRY, L., & TEHAN, G. (2005). An item/order trade-off explanation of word length and generation effects. *Memory*, 13, 364-371.
- HULME, C., SURPRENANT, A. M., BIRETA, T. J., STUART, G., & NEATH, I. (2004). Abolishing the word-length effect. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 30, 98-106.
- LA POINTE, L. B., & ENGLE, R. W. (1990). Simple and complex word spans as measures of working memory capacity. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 16, 1118-1133.
- LOVATT, P., AVONS, S. E., & MASTERSON, J. (2000). The word-length effect and disyllabic words. *Quarterly Journal of Experimental Psychology*, 53A, 1-22.
- LOVATT, P., AVONS, S. E., & MASTERSON, J. (2002). Output decay in immediate serial recall: Speech time revisited. *Journal of Memory & Language*, 46, 227-243.
- MUELLER, S. T., SEYMOUR, T. L., KIERAS, D. E., & MEYER, D. E. (2003). Theoretical implications of articulatory duration, phonological similarity and phonological complexity in verbal working memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 29, 1353-1380.
- 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. (2004). Separating item and order information through process dissociation. *Journal of Memory & Language*, 50, 113-133.
- NAIRNE, J. S., NEATH, I., & SERRA, M. (1997). Proactive interference plays a role in the word-length effect. *Psychonomic Bulletin & Review*, 4, 541-545.

- NEATH, I., BIRETA, T. J., & SURPRENANT, A. M. (2003). The time-based word length effect and stimulus set specificity. *Psychonomic Bulletin & Review*, 10, 430-434.
- NEATH, I., & NAIRNE, J. S. (1995). Word-length effects in immediate memory: Overwriting trace decay theory. *Psychonomic Bulletin & Review*, 2, 429-441.
- POIRIER, M., & SAINT-AUBIN, J. (1995). Memory for related and unrelated words: Further evidence on the influence of semantic factors in immediate serial recall. *Quarterly Journal of Experimental Psychology*, 48A, 384-404.
- 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, & Cognition*, 28, 1019-1034.
- SAINT-AUBIN, J., & POIRIER, M. (1999). Semantic similarity and immediate serial recall: Is there a detrimental effect on order information? *Quarterly Journal of Experimental Psychology*, 52A, 367-394.
- SCHWEICKERT, R. (1993). A multinomial processing tree model for degradation and redintegration in immediate recall. *Memory & Cognition*, 21, 168-175.
- SERVICE, E. (1998). The effect of word length on immediate serial recall depends on phonological complexity, not articulatory duration. *Quarterly Journal of Experimental Psychology*, 51A, 283-304.
- TEHAN, G., & HUMPHREYS, M. S. (1996). Cuing effects in short-term recall. *Memory & Cognition*, 24, 719-732.
- TEHAN, G., & TURCOTTE, J. (2002). Word-length effects are not due to proactive interference. *Memory*, 10, 139-149.
- TOLAN, G. A., & TEHAN, G. (2005). Is spoken duration a sufficient explanation of the word length effect? *Memory*, 13, 372-379.
- WICKENS, D. D. (1970). Encoding categories of words: An empirical approach to meaning. *Psychological Review*, 77, 1-15.

NOTES

1. This could be due to differential word familiarity; homogeneous lists and heterogeneous lists were counterbalanced, but different words were used in homogeneous blocks, due to the restrictions mentioned earlier. Post hoc familiarity ratings on a 7-point scale for the entire word pool from 50 new participants revealed no reliable differences among the word lists, ruling out this explanation.
2. Analyses without the homogeneous list conditions, whose inclusion may have affected the results, due to the lack of protrusion errors, revealed identical results.
3. Length $\times$ trial analyses with homogeneous blocks and heterogeneous lists revealed no interactions ($F_s < 1.1$), but significant linear trends for trial, showing more protrusions in the latter trials [$F_s(1,49) > 10.8$, $MS_{\epsilon}s < 0.004$]. Hence, there was direct evidence of PI buildup in these conditions, but no evidence that the WLE's magnitude increased with it.
4. The possibility that the WLE attenuation is due to increased rehearsal speed when words are from the same category was ruled out post hoc. Seventeen new participants were asked to quickly rehearse five-word lists randomly selected from the word pool in the homogeneous and heterogeneous conditions. No similarity effects were observed for their average rehearsal times.
5. Main effects of similarity for transpositions and item errors [$F_s(2,368) > 4.1$, $MS_{\epsilon}s < 0.004$] were such that heterogeneous lists had the most item errors, whereas homogeneous lists had the most transpositions. The length $\times$ similarity interaction for omissions [$F(2,368) = 4.32$, $MS_{\epsilon} = 0.01$] was due to reliably fewer short-word omissions with homogeneous blocks than with heterogeneous lists [$t(122) = 2.52$] and reliable increases in long-word omissions when going from homogeneous lists to homogeneous blocks to heterogeneous lists ($t_s > 1.9$).

(Manuscript received June 11, 2004;
revision accepted for publication April 8, 2006.)