

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/7062325>

The Distinctiveness of the Word–Length Effect.

Article in *Journal of Experimental Psychology Learning Memory and Cognition* · June 2006

DOI: 10.1037/0278-7393.32.3.586 · Source: PubMed

CITATIONS

27

READS

916

6 authors, including:



George Stuart

Central Queensland University

17 PUBLICATIONS 1,420 CITATIONS

[SEE PROFILE](#)



Aimee Surprenant

Memorial University of Newfoundland

93 PUBLICATIONS 1,433 CITATIONS

[SEE PROFILE](#)



Gordon Brown

The University of Warwick

142 PUBLICATIONS 8,677 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



Word length effect [View project](#)

The Distinctiveness of the Word-Length Effect

Charles Hulme
University of York

Ian Neath
Purdue University

George Stuart
Goldsmiths College, University of London

Lisa Shostak
University of York

Aimée M. Surprenant
Purdue University

Gordon D. A. Brown
University of Warwick

The authors report 2 experiments that compare the serial recall of pure lists of long words, pure lists of short words, and lists of long or short words containing just a single isolated word of a different length. In both experiments for pure lists, there was a substantial recall advantage for short words; the isolated words were recalled better than other words in the same list, and there was a reverse word-length effect: Isolated long words were recalled better than isolated short words. These results contradict models that seek to explain the word-length effect in terms of list-based accounts of rehearsal speed or in terms of item-based effects (such as difficulty of assembling items).

Keywords: serial recall, word-length effect, distinctiveness

Immediate serial recall, repeating a list of words immediately after hearing it, is a seemingly simple task, but despite more than a century of study, we still lack a generally accepted theoretical account of how we perform this task (Nairne, 2002). One theoretically influential and much studied phenomenon is the *word-length effect*: the finding that lists of short words are substantially better recalled than lists of long words (Baddeley, Thomson, & Buchanan, 1975). Current theoretical accounts of the word-length effect can be divided into two very different forms of explanation: primarily item and list based (see Hulme, Surprenant, Bireta, Stuart, & Neath, 2004). According to item-based explanations, short words are inherently easier to recall than long words (perhaps because they contain fewer segments or are in some other way simpler, e.g., Neath & Nairne, 1995). In contrast, according to list-based explanations, the word-length effect is the product of some overall characteristic of the list (typically, the overall spoken duration of the list, which determines how easily the list can be maintained in memory by covert rehearsal, e.g., Baddeley et al., 1975; Burgess & Hitch, 1999).

Hulme et al. (2004) tested the contrasting predictions made by item- and list-based explanations of the word-length effect and produced data that were incompatible with both forms of explanation. That study compared the recall of pure lists of long words (five syllables), pure lists of short words (one syllable), and mixed lists of equal proportions of both long and short words. In pure lists, long words were much more poorly recalled than short words (the typical word-length effect). However, in mixed lists, the word-length effect was abolished, and both long and short words were recalled as well as short words in pure lists.

This pattern of results refutes predictions derived from item-based models, such as the feature model of Neath and Nairne (1995) or the model of Brown and Hulme (1995), which predicts a “sawtooth” serial position function for alternating lists of long and short words. The results also refute predictions from list-based models, which predict smooth serial position curves for alternating lists of long and short words, which should be at intermediate levels compared with the recall functions for pure lists of long or short words (e.g., Burgess & Hitch, 1999). In general terms, if the ease with which words are recalled depends upon the time required to rehearse the list, then lists that contain equal numbers of long and short words should be recalled at intermediate levels compared with pure lists of short or long words.

On the basis of these results, Hulme et al. (2004) proposed a novel explanation of the word-length effect, drawing on ideas of item complexity and item distinctiveness. They argued that the word-length effect depends upon the phonological complexity of items and not upon their spoken length per se. However, they argued that, in addition to the role of item complexity, how well any given item in a list can be retrieved will depend upon the relative distinctiveness of the item in relation to other items in the list. In this view, the recall of long items in mixed lists is boosted

Charles Hulme and Lisa Shostak, Department of Psychology, University of York, York, England; Ian Neath and Aimée M. Surprenant, Department of Psychology, Purdue University; George Stuart, Department of Psychology, Goldsmith College, University of London, London, England; and Gordon D. A. Brown, Department of Psychology, University of Warwick, Coventry, England.

We thank Philip Quinlan for statistical advice and Nelson Cowan, Elisabet Service, and Gerry Tehan for helpful comments on a version of this article.

Correspondence concerning this article should be addressed to Charles Hulme, Department of Psychology, University of York, York YO10 5DD, England. E-mail: ch1@york.ac.uk

by their relatively greater distinctiveness in mixed lists than in pure lists. In support of this explanation, the data from these experiments were well fitted by a computer simulation that embodies item distinctiveness as a critical determinant of recall (SIMPLE; Brown, Neath, & Chater, 2002; Neath & Brown, 2006).

Cowan, Baddeley, Elliott, and Norris (2003) performed a conceptually similar experiment in which they varied the number of long items in a six-item list from zero (pure short) to six (pure long). In the case in which they had three short and three long items, they found a significant word-length effect, contrary to what Hulme et al. (2004) reported. Bireta, Neath, and Surprenant (in press) explored these apparently contradictory findings and concluded that the difference was due to the particular stimulus sets used: The same six short and six long words used by Cowan et al. yielded results that replicated their findings, whereas three other sets of stimuli yielded results that replicated the findings of Hulme et al. Bireta et al. speculated that the particular sets of six short and six long words chosen by Cowan et al. differed systematically in ways other than length. (The short words used by Cowan et al. appeared to be more concrete and imageable than the long words.) It appears therefore that the pattern reported by Hulme et al. is the more general and that, when equal numbers of short and long words are presented in the same list, the word-length effect is abolished, with both types of words being recalled at levels indistinguishable from short words in pure lists. This is a surprising pattern that appears to refute current list- and item-based accounts of the word-length effect (Bireta et al., in press; Hulme et al., 2004) and that provides support for the importance of distinctiveness as a critical determinant of the word-length effect and serial recall tasks more generally.

The aim of this study was to test the distinctiveness account of the word-length effect and its abolition in mixed lists together with the contrasting predictions from list- and item-based theories. What should happen if we present a list of six words containing five short words and one long word? According to item-based accounts (such as the feature model; Neath & Nairne, 1995), the isolated long word should not benefit from being in a list of short words. On the other hand, according to list-based explanations, there should be a fairly large advantage for the single long word, so it should be recalled almost as well as a short word in a pure list of short words, and this in turn should go hand in hand with a small decrement in recall of the five short words in the list. Graphically, this would be a smooth serial position curve, with the mixed list (five short words and one long word) being recalled slightly less well than a pure list of short words.

In contrast, the distinctiveness explanation proposed by Hulme et al. (2004) makes the somewhat counterintuitive prediction that the isolated long word in a list of five short words should actually be recalled *better* than a comparable short word in a pure list of short words. The reason is that the long word in such a list benefits from being highly distinctive from other words in the list. Such a pattern is clearly incompatible with predictions from the list- or item-based accounts of the word-length effect just described. The account also predicts an equal advantage for an isolated short item.

Cowan et al. (2003) included a condition in which there was an isolated long word among short words and a condition in which an isolated short word was placed among long words. They found that the proportion of isolated long words correctly recalled was 0.66 compared with 0.61 for short words in pure short lists. The

proportion of isolated short words correctly recalled was 0.61, the same value as for short words in pure lists. Although some aspects of the results of Cowan et al. were judged by Bireta et al. (in press) to be due to the particular stimulus set used, this pattern was replicated by Bireta et al. with a different set of stimuli. However, in both of these studies, the position of the isolated item was determined randomly, so it did not appear equally often in each serial position. It is thus unclear what the resulting serial position curves would be like.

The two experiments reported here were designed to test a specific prediction derived from the version of SIMPLE that Hulme et al. (2004) used to explain their results. According to this model, isolated short and long words should both be recalled better than short words from pure lists, and both types of isolated words should be recalled equally well. In addition, the data will allow us to examine the isolation effect at each serial position, which has not been possible in previous studies.

Experiment 1

To test these contrasting predictions here, we compared the immediate serial recall of pure lists of long words, pure lists of short words, and lists of long or short words containing just a single isolated word of a different length. On the basis of the notion of distinctiveness being crucial for retrieval, we expected the isolated words to benefit at recall compared with words of the same length in pure lists. Specifically, we expected that the isolated words (whether long or short) would be recalled better than the remaining "background" items in the mixed lists and the short words in the pure lists.

Method

Participants. Sixteen undergraduate students from the University of York participated. The sample comprised 9 men and 7 women, and the age range was 19–24 years. All participants were native speakers of English.

Stimuli. A set of eight 5-syllable words (*aluminum, hippopotamus, periodical, physiology, refrigerator, tuberculosis, university, Yugoslavia*) and eight 1-syllable words (*zinc, stoat, scroll, maths, switch, mumps, school, Greece*) was used. These words (Hulme, Maughan, & Brown, 1991; Hulme et al., 2004) are a subset of the words originally used by Baddeley et al. (1975) and are matched individually for frequency and conceptual class. The words were recorded by a female speaker, digitized at a sampling rate of 22 kHz, and stored on a Macintosh computer. Custom-built software was used to control the order and rate of presentation of these stimuli via an amplified speaker.

Design and procedure. To familiarize participants with the words used in the experiment, they were first asked to listen to each of the 16 spoken words in turn. If there was any uncertainty in their perception or understanding of a word, its meaning was explained, and the word was played once more.

Serial recall was then tested over a series of 96 trials divided into two blocks of 48 trials. One block of 48 trials consisted of 24 lists of six short words (pure short) and 24 lists of five short words with one isolated long word (short background, isolated long). The other block of 48 trials consisted of 24 lists of six long words (pure long) and 24 lists of five long words with one isolated short word (long background, isolated short). The isolated long or short words occurred equally often in each of the six serial positions. The six items presented in each trial were selected randomly without replacement from one or both of the eight-item pools and were presented at a rate of one word/s. In both blocks, trial types (pure or isolated) occurred randomly. Order of blocks was counterbalanced across

participants. At the end of a list, participants attempted to recall it aloud in the order of presentation. If participants were unable to recall an item from a particular serial position, they were asked to say "blank" and to continue with the next item in the series.

Results and Discussion

Correct recall. Each response was counted as correct if it was the correct word recalled in the correct position. The most informative way to analyze the data for the isolated conditions is to derive composite lists: short background, isolated long; and long background, isolated short. This allows direct comparisons to be made between how well a word of a given length (short or long) is recalled as a function of the list type it is embedded within (pure short, pure long, isolated short, isolated long, short background, long background). The data in Figure 1 are plotted in this way as a function of serial position, and the data in Figure 2 show the mean levels of recall in each context (and 95% within-subjects confidence intervals; Loftus & Masson, 1994; Masson & Loftus, 2003), ignoring serial position.

The data for the pure lists show a large advantage for the recall of short words, replicating earlier findings (e.g., Baddeley et al., 1975; Cowan et al., 1992; Hulme et al., 2004). The short and long background items appear to be recalled at similar levels compared with the corresponding pure lists. However, most strikingly, the isolated long words are recalled better than the isolated short words (a reverse word-length effect), and both are recalled better than the short words in pure lists.

A three-way analysis of variance (ANOVA) of the data shown in Figure 1 with list type (pure, isolated, or background), length

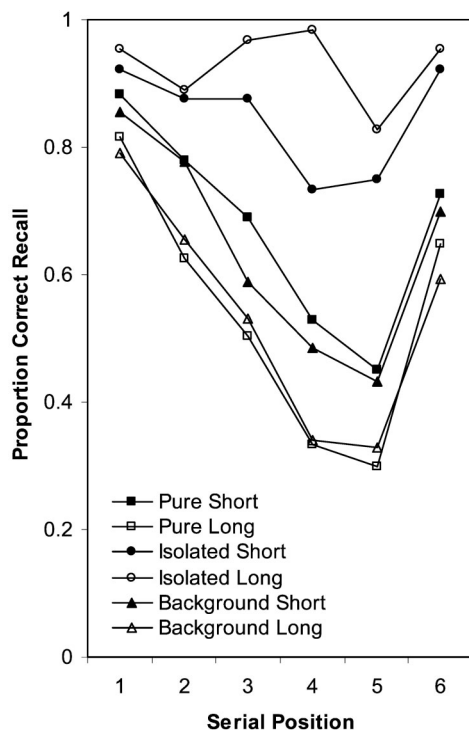


Figure 1. Mean proportion of words recalled in the correct position as a function of list type and serial position in Experiment 1.

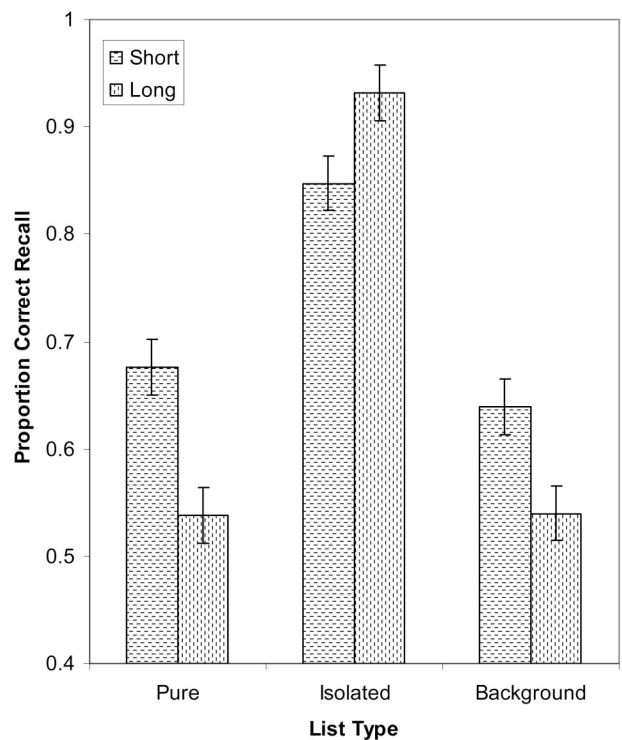


Figure 2. Mean proportion (and 95% within-subjects confidence intervals) of words recalled in the correct position as a function of list type in Experiment 1.

(short or long), and serial position (1–6) as within-subjects variables revealed statistically significant main effects of list type, $F(2, 30) = 150.90$ ($MSE = 0.036$); serial position, $F(5, 75) = 34.12$ ($MSE = 1.70$); and length $F(1, 15) = 6.21$ ($MSE = 0.385$). There were also statistically significant interactions between list type and length, $F(2, 30) = 44.13$ ($MSE = 0.015$); list type and position, $F(10, 150) = 12.40$ ($MSE = 0.016$); and List Type $\times$ Length $\times$ Position, $F(10, 150) = 2.67$ ($MSE = 0.042$).

It is clear from Figure 2 that, although recall of words in the pure and background lists is at essentially equivalent levels, isolated lists are recalled better and show a reverse word-length effect (better recall of long words in the isolated lists). To investigate this pattern further, two additional ANOVAs were run: one comparing pure lists with isolated lists and one comparing pure lists with background lists.

An ANOVA comparing performance on the pure and background lists in which the variables were list type (pure and background), length (short and long), and serial position (1–6) revealed significant main effects of length, $F(1, 15) = 23.03$ ($MSE = 0.059$), and serial position, $F(5, 75) = 44.40$ ($MSE = 0.044$). Most critically, however, the main effect of list type was not significant, $F(1, 15) = 1.63$ ($MSE = 0.017$), and there were no significant interactions with list type. This confirms what is shown in Figure 2: Background lists of long and short words are recalled at very similar levels compared with the pure lists.

An equivalent ANOVA comparing pure and isolated lists revealed significant main effects of list type, $F(1, 15) = 164.02$ ($MSE = 0.046$), and serial position, $F(5, 75) = 24.17$ ($MSE =$

0.035), but no significant main effect of length, $F(1, 15) = 1.74$ ($MSE = 0.074$). Most critically, however, there was a highly significant interaction between list type and length, $F(1, 15) = 95.02$ ($MSE = 0.012$). This confirms the pattern shown in Figure 2: For the pure lists, short words are recalled better than long words, whereas for the isolated items, long words are recalled better than short words. There was also a significant interaction between list type and position, $F(5, 75) = 15.156$ ($MSE = 0.019$), which reflects the fact that the serial position curves for the pure lists were more deeply bowed than the functions for the isolated words. (This pattern may simply be a product of the very high levels of accuracy in the recall of the isolated words.)

Pattern of recall errors. The pattern of recall errors was analyzed to see what further light it might throw on the processes operating. For clarity, the composite lists were used for this analysis also. Order errors are errors in which an item that had been presented in the list was recalled but in the wrong position. Item errors are errors in which participants failed to recall the correct item. Item errors can be broken down into three subtypes: omissions, in which participants responded "blank" or indicated in another way that they did not know a given item; intraset intrusions, in which an item recalled was from the set of stimuli for that condition; and extraset intrusions, in which a nonexperimental word was recalled.

The error rates, along with the proportion of items correct in each condition, are shown in Table 1. The pattern of errors shown in Table 1 indicates that performance on pure and background lists of short and long words is nearly identical in terms of the pattern of recall errors. There are no selective differences in order errors, and the differences in recall accuracy between long and short words appear to be attributable to differences in rates of omission errors. (This is the same pattern reported by Hulme et al., 2004.) Similarly, comparison of performance on the isolated long words and the short background items and of the isolated short words and the long background items shows that improvements in recall on the isolated words appear to reflect large reductions in the rates of omission errors for the isolated words. As in Hulme et al., virtually no extraset intrusions were made; just 2 participants made one extraset intrusion (to a pure short-word list).

Summary. The results of Experiment 1 partially support the predictions made by Hulme et al. (2004): The isolated long words were indeed recalled better than the short words from pure lists. However, the isolated long words were also recalled better than the

isolated short words, whereas the model predicts equivalent recall for these items. Given that only one small set of words was used here, Experiment 2 was designed to explore the generality of these effects by using a different larger set of words and a different modality of presentation and response.

Experiment 2

The pattern of results presented in Experiment 1 was extremely clear. Most strikingly, there was a large advantage for the recall of short words compared with long words in pure lists, but this effect was reversed for isolated words presented in lists of a different background length. So, when a single long word occurred in a list in which the other items were five short words, the isolated long word was actually recalled much better than a corresponding short word in a pure list of short words and also better than a single short word occurring in a list in which the other five items were long words. The recall of isolated words resulted in greatly enhanced recall performance coupled with a reverse word-length effect.

Experiment 1 used a small closed pool of items, as is typical in studies of immediate serial recall (e.g., Baddeley et al., 1975; Cowan et al., 2003; Hulme et al., 1991). Arguably, however, this might place limits on the generality of the effect (Bireta et al., in press; Neath, Bireta, & Surprenant, 2003). It is possible that the striking enhancement of recall for isolated items coupled with a reversal of the word-length effect found here would not occur if a larger, and so less familiar, open word set were used. We decided therefore to replicate this effect in Experiment 2 using a pool of 80 short and 80 long items. In this experiment, we also switched to visual rather than auditory presentation with a serial reconstruction of order task as the mode of response: All words presented in a list appeared on the screen at the end of a trial, and participants arranged them in the order of presentation. This change in response mode has the added advantage of effectively equating output times for long- and short-word lists. One explanation of the word-length effect in studies using spoken recall has been in terms of output decay. According to this view, one major source of forgetting is decay during output, and given that long words take longer to say at recall than short words, at least part of the word-length effect may arise from output delays (e.g., Cowan et al., 1992). The use of an order reconstruction procedure here equates output times for short and long words and thus eliminates one source of output

Table 1
Proportion of Items Correct and Proportion of Error Types as a Function of List Type

List type	Correct	Error type				
		Order errors	Item errors			
			Total	Omissions	Intraset intrusions	Extraset intrusions
Pure short	0.676	0.146	0.176	0.152	0.024	0.001
Pure long	0.541	0.149	0.309	0.288	0.021	0.000
Short background	0.640	0.146	0.214	0.184	0.030	0.000
Long background	0.540	0.150	0.308	0.291	0.017	0.000
Isolated short	0.846	0.055	0.102	0.102	0.000	0.000
Isolated long	0.930	0.042	0.028	0.023	0.005	0.000

effects as a potential explanation of any word-length effect obtained (cf. Neath et al., 2003).

The method used in this experiment is essentially a modification of that used in Experiment 2 of our earlier study (Hulme et al., 2004), in which we showed that presenting alternating short and long words in the same list abolished the word-length effect using the order reconstruction procedure, just as it had done in an immediate verbal serial recall task.

Method

Participants. Fifty Purdue University undergraduate students participated in exchange for credit in introductory psychology courses. All participants were native speakers of American English and were between 18 and 23 years old ($M = 18.92$); there were 20 men and 30 women.

Stimuli. Eighty short (one syllable) and 80 long (three to five syllables) words from LaPointe and Engle (1990) were used. These stimuli have previously been shown to produce robust word-length effects (e.g., Hulme et al., 2004, Experiment 2).

Design and procedure. There were 14 list types. One type of list had only short words (pure short), and a second had only long words (pure long). Six list types had only one short and five long words, with the short word occurring at each serial position. The remaining six list types had only one long and five short words, with the long word occurring at each serial position. Thus, each participant received a total of 112 lists, divided into 8 lists of the 14 types. The different list types occurred in random order, and items in each list type were drawn randomly without replacement from the sets of 80 short and 80 long words.

Each word was shown in 24-point Helvetica lowercase type for 1 s in black against a white background; after the final word had been shown, the phrase *Recall Words* appeared, and six buttons appeared (two rows of three buttons on the right side of the screen) labeled with the six words from that trial in alphabetical order. The participant was asked to click on the buttons to recreate the presentation order; once a response had been made, it could not be changed, as the button was deactivated. Omissions were not possible. Recall was self-paced. After six responses had been made, a button labeled *Next Trial* became active, and the next trial began when the participant clicked on this button. Participants were tested individually, and an experimenter remained in the room to ensure compliance with the instructions.

Results and Discussion

The serial position functions for this experiment are plotted in Figure 3, and the mean levels of recall (and 95% within-subjects confidence intervals; Loftus & Masson, 1994; Masson & Loftus, 2003) are shown in Figure 4.

A direct comparison between Figures 2 and 4 is particularly revealing in showing the similarity in the results between Experiments 1 and 2. In both experiments for pure lists, there is a substantial recall advantage for short words (a 14% word-length effect in both experiments). Similarly, in both experiments, there is an advantage for the recall of background short words compared with background long words that is similar in size to the word-length effect for pure lists (a 10% word-length effect for background items in Experiment 1 and an 11% word-length effect for background items in Experiment 2). Finally and most critically, the word-length effect is reversed for the isolated items in both experiments (an 8% reverse word-length effect for isolated items in Experiment 1 and a 9% reverse word-length effect for isolated items in Experiment 2).

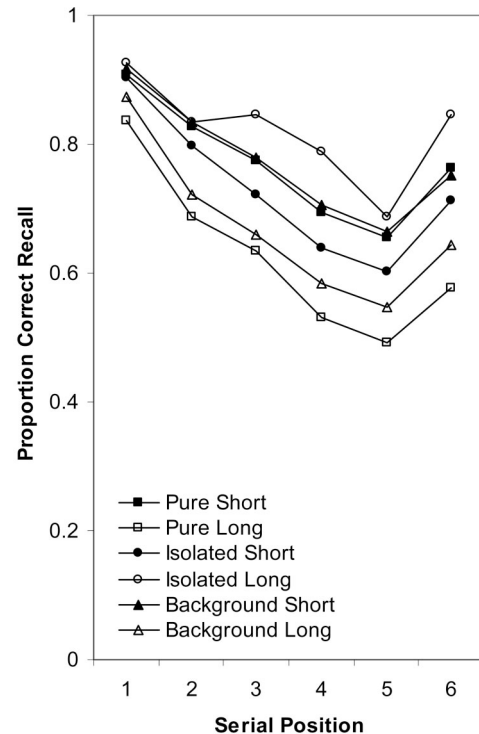


Figure 3. Mean proportion of words recalled in the correct position as a function of list type and serial position in Experiment 2.

A three-way ANOVA of the data shown in Figure 3 with list type (pure, isolated, or background), length (short or long), and serial position (1–6) as within-subjects variables revealed statistically significant main effects of list type, $F(2, 98) = 49.28$ ($MSE = 0.018$); serial position, $F(5, 245) = 87.37$ ($MSE = 0.0345$); and length, $F(1, 49) = 26.73$ ($MSE = 0.0452$). There were also statistically significant interactions between list type and length, $F(2, 98) = 51.09$ ($MSE = 0.046$), and List Type $\times$ Length $\times$ Position, $F(10, 490) = 3.75$ ($MSE = 0.014$).

It is clear from Figure 4 that, although recall of words in the pure and background lists is at similar levels, with a substantial advantage for short words, the word-length effect is reversed for the isolated words (as it was in Experiment 1; see Figure 2). To investigate this pattern further, two additional ANOVAs were run: one comparing the pure lists with isolated lists and one comparing the pure lists with background lists.

An ANOVA comparing performance on the pure lists and the background lists in which the variables were list type (pure and background), length (short and long), and serial position (1–6) revealed significant main effects of list type, $F(1, 49) = 9.58$ ($MSE = 0.019$); length, $F(1, 49) = 62.58$ ($MSE = 0.073$); and serial position, $F(5, 245) = 101.08$ ($MSE = 0.021$). The effect of list type here confirms (as shown in Figure 4) that the background lists were recalled slightly better than the pure lists. However, there were no significant interactions with list type, and there was a substantial word-length effect in both the pure and background lists. Finally, there was a significant interaction between length and position, $F(5, 245) = 4.75$ ($MSE = 0.021$), which reflects the slightly flatter serial position functions for the short-word lists

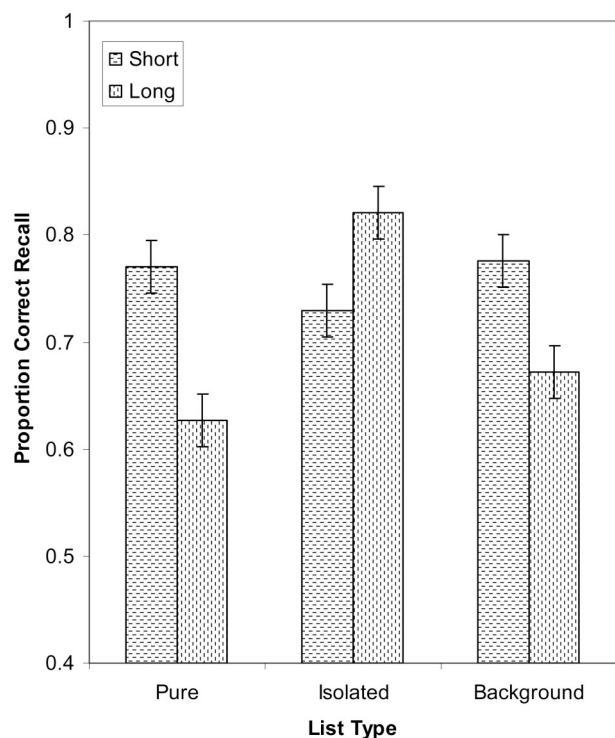


Figure 4. Mean proportion (and 95% within-subjects confidence intervals) of words recalled in the correct position as a function of list type in Experiment 2.

(which may simply reflect the higher levels of performance on those lists).

An equivalent ANOVA comparing pure with isolated lists revealed significant main effects of list type, $F(1, 49) = 61.96$ ($MSE = 0.028$), and serial position, $F(5, 245) = 101.08$ ($MSE = 0.021$), but no significant main effect of length, $F(1, 49) = 3.76$ ($MSE = 0.054$). Most critically, however, there was a highly significant interaction between list type and length, $F(1, 49) = 61.67$ ($MSE = 0.067$), which reflects the fact that the word-length effect was reversed for the isolated words. The only other significant effect in this analysis was the three-way interaction between list type, length, and serial position, $F(5, 245) = 4.94$ ($MSE = 0.018$). This appears to reflect the differently shaped serial position function for the isolated long words, which shows a particularly marked recall advantage in later serial positions.

As noted above, in most major respects, the results from Experiments 1 and 2 are very similar, and most critically, the word-length effect is reversed for the isolated words in both experiments. There are a couple of ways in which the results from the two experiments differ, however. First, for the pure and background lists, it is clear that performance is better in Experiment 2 than in Experiment 1. This is most easily explained in terms of the different response requirements of the two experiments. Arguably, the order reconstruction procedure used in Experiment 2 is a less demanding task (that minimizes the demands on item memory) than the serial recall task used in Experiment 1.

Second, the isolated words in Experiment 2 are recalled at levels more similar to the levels of words recalled in pure lists in

Experiment 2. (As shown in Figure 4, the levels of recall for pure and isolated short words do not differ, whereas the isolated long words are recalled better than long words in the pure lists.) In contrast, in Experiment 1, the isolated long and short words are recalled substantially better than the short and long words in pure lists. The most likely explanation for this discrepancy is the differences between the stimulus set characteristics in the two experiments. In Experiment 1, the short words were all one syllable, and the long words were all five syllables. In Experiment 2, however, the difference between short and long words was much less stark: The short words were all one syllable, but the long words could contain three, four, or five syllables. This difference can arguably explain the larger advantage shown by the isolated words in Experiment 1 because the difference in length between the isolated words and background items was larger in Experiment 1 than in Experiment 2.

General Discussion

In these experiments, we compared immediate serial recall of short and long words in six contexts: short words in pure lists, long words in pure lists, short words in background lists, long words in background lists, short words in “isolated” lists, and long words in isolated lists. The data for the pure lists show a large advantage for the recall of short words, replicating earlier findings (e.g., Baddeley et al., 1975; Cowan et al., 1992; Hulme et al., 2004). The short and long background items in the lists with isolated words are recalled at similar levels compared with the corresponding pure lists. In Experiment 1, both types of isolated words are recalled much better than the background items in the same list, and the isolated long words are actually recalled better than the isolated short words (a reverse word-length effect).

In Experiment 2, the results are very similar with one important difference: Recall of the isolated short items was not different from that of the pure short items, even though we did replicate the reverse word-length effect in the isolated long condition. We explain this disparity in terms of the relative distinctiveness of the isolated item within the list context: The isolated long word is an outlier in the list (one long word in the context of five 1-syllable words), whereas the isolated short word occurs in the context of greater variability within the list (one short word in the context of five words that may include three-, four-, and five-syllable words in varying numbers).

The improvements in recall of isolated items seen in both experiments provide a further demonstration of the powerful effects of isolation on memory first described by von Restorff (1933). In her classic experiments, she found that a single isolated item (lists of nine numbers and one nonsense syllable or nine nonsense syllables and one number) was recalled substantially better than the background items. Such isolation effects have been replicated many times with a wide variety of materials and procedures (for reviews, see Kelley & Nairne, 2001; Schmidt, 1991). Kelley and Nairne showed that isolating a word (by changing the font size) improved recall of the item in an order reconstruction task that was equivalent to the procedure used in Experiment 2. They also found an equivalent effect when participants were required to generate one item in the list (using word fragment completion). The explanation offered by Kelley and Nairne for these isolation effects is similar to the approach adopted here.

Specifically, they related the effect of isolation to the feature model (Nairne, 1990) and postulated that the isolated item benefited at retrieval from its distinctiveness from other items in a memory search set. The version of SIMPLE proposed to account for word-length effects (Hulme et al., 2004) is related to the feature model insofar as both postulate that one critical determinant of retrieval is the relative distinctiveness of item representations.

The reverse word-length effect found here for isolated items (the better recall of isolated long words than isolated short words) is reminiscent of the reverse word-length effect found by Hendry and Tehan (2005) when they assessed delayed recognition of short and long words. Hendry and Tehan explained this reverse word-length effect in terms of trade-offs between the encoding of item and order information. In this view, long items require greater processing at encoding because of their greater length and complexity, and this greater processing confers advantages for item memory at the expense of worse order memory. (Order memory for long words suffers due to the greater attentional resources required for encoding item information for these items.)

Such an explanation cannot account for the pattern of data reported here, however. Hendry and Tehan (2005) noted that "the item/order approach does predict that the standard word length effect should be apparent in any task that requires the use of serial order information," and both present experiments involved serial order tasks. The isolated words in Experiment 1 were recalled in correct position substantially better than the pure or background items of comparable length, but this serial recall advantage was not produced at the expense of worse order memory for the background items in the same lists because those items were recalled at essentially identical levels compared with the items of corresponding length in pure lists. Furthermore, although the effects obtained here in both experiments involved changes in the accuracy of serial memory (requiring both item and order information), an analysis of the recall errors in Experiment 1 demonstrated that the improvements in recall of isolated words appear to be mediated by improvements in item memory (a reduction in omissions) in the absence of any substantial change in order memory for other list items. Finally, a prediction of the item/order trade-off account of the word-length effect is that, for pure lists, differences in recall between long and short words should reflect differences in order memory. However, in Experiment 1, differences in recall between pure lists of long and short words reflected differences in omission errors with essentially identical rates of order errors (the same pattern as reported previously by Hulme et al., 2004). Therefore, we believe the present data raise serious problems for attempts to account for word-length effects in terms of trade-offs between the encoding of item and order information.

The pattern of results reported here has important implications for theories of the word-length effect and for models of serial recall more generally. Our results are clearly at odds with the phonological loop model (Baddeley, 1986; Burgess & Hitch, 1999), which postulates that the word-length effect arises from decay that can be offset by rehearsal. Word length on this account is claimed to affect the rate at which items can be rehearsed, and such a view leads to the prediction that inserting a single long word into a list of short words should simply reduce the level of recall for all words in the list. The present results refute this prediction.

On the other hand, item-based theories of the word-length effect (e.g., Brown & Hulme, 1995; Neath & Nairne, 1995) predict that

an isolated word that differs in length from other items in a list should simply be recalled at the same level as in a pure list. (So, recall of an isolated word should be neither benefited nor harmed by being in a list of items of different length.) It is clear from the results presented here, however, that the recall of isolated words in mixed lists (whether long or short) enjoys a substantial advantage compared with how those same items would be recalled in pure lists in which all words are of uniform length. These findings therefore contradict the feature model (Neath & Nairne) and other item-based theories of the word-length effect (e.g., Brown & Hulme).

The basic pattern for lists with an isolated word of different length obtained here is, however, broadly in line with the predictions from the distinctiveness account of the word-length effect derived from the SIMPLE model (Brown et al., 2002) and advanced by Hulme et al. (2004). Any account of the word-length effect needs to explain two key recent results: (a) the dramatic improvement in recall of long items in mixed lists reported by Hulme et al. and (b) the finding that isolated long items are recalled better than isolated short items reported here.

Hulme et al. (2004) argued that the first finding eliminates all extant accounts of the word-length effect. Models based on the temporal duration of the list of items (e.g., Burgess & Hitch, 1999) predict that performance on mixed lists will be intermediate between that on pure short and pure long lists because the duration of the mixed lists will be, on average, longer than that of pure short lists (because there are some long words) but shorter than that of pure long lists (because there are some short words). Models based on explanations at the level of the item (e.g., Brown & Hulme, 1995; Neath & Nairne, 1995) predict that word-length effects will be obtained in mixed lists. Both of these predictions are incorrect.

One model that is able to account for the results from the mixed lists (Hulme et al., 2004) is an extension of SIMPLE (Brown et al., 2002). By this account, word-length effects are due to item complexity (see, e.g., Caplan, Rochon, & Waters, 1992; Neath et al., 2003; Service, 1998), and this complexity affects the relative discriminability of the items in the list. Short items are, on average, more easily discriminable than long items. In lists that consist of mixtures of long and short words, the relative distinctiveness of the long items will be greatly increased. Intuitively, this can be appreciated by considering the difference between two lists: a pure list of long words (e.g., *physiology*, *tuberculosis*, *periodical*, *refrigerator*, *aluminum*, *hippopotamus*) and a list of both short and long items (e.g., *maths*, *tuberculosis*, *scroll*, *refrigerator*, *zinc*, *hippopotamus*). The long items in the second list stand out more than the comparable items in the pure list.

To account for the abolition of the word-length effect that they found in alternating lists of words, Hulme et al. (2004) presented a simulation of their data using SIMPLE. In that simulation, the words in the to-be-remembered lists were represented in memory along three dimensions. The temporal dimension represents the time of item presentation. The item dimension represents the relative discriminability of items. (It was assumed that the representations of short items are, on average, more easily discriminable from one another than the representations of long items.) One way of thinking about this assumption within the model is that short words, being less complex, are encoded in a more efficient and complete manner than long words, which are more complex. The class dimension represents the emergent difference that becomes

apparent when short and long words are combined in a list (to capture the idea that *stoat* and *hippopotamus* are perceived as radically different in length when placed together in adjacent positions in a list). In the simulations presented by Hulme et al., the recall advantage for short words compared with long words in pure lists was accounted for purely in terms of the item dimension: Short words were simply better represented (with more discriminable representations) than long words. In the alternating lists, however, in which short and long words co-occurred, the class dimension produced nonzero values, and the effect of this in the simulations presented was to boost the recall of long items to the same level as short words (because the long items in alternating lists became just as discriminable as the short items because of their higher values on the class dimension).

The simulation presented by Hulme et al. (2004) has been applied to the present results. The results of that simulation showed that isolated words benefit compared with items in pure lists. The reason is that, in addition to the emergent distinctiveness arising from the class dimension, the item is likely to be represented in areas of psychological space with few neighbors (because all of the other items are of the other type). Thus, both isolated long and short items benefit from isolation. However, the simulation produces essentially identical levels of recall for isolated long and short words and hence cannot account for the reverse word-length effect obtained here. Intuitively, this can be understood as due to symmetry in the way differences are calculated (i.e., the long items are as different from short items as the short items are from long items).

Some initial explorations of alternative simulation strategies within SIMPLE suggest that, with modifications to the coding scheme used, the model may be able to account for the results. For example, a dimension representing phonological information might produce the asymmetrical isolate data. With five long items and one short item, there are more likely to be more phonemes shared between the background items and the isolated word than with five short items and one long item.¹ An alternative explanation might arise from an asymmetry in the relative number of short and long words in the vocabulary. Because there are currently no data to constrain the choice of modification, we choose to leave the model making one incorrect prediction (predicting that the recall of isolated short words should be equivalent to, not worse than, the recall of isolated long words) but succeeding in all other respects.

We argue, however, that two important conclusions concerning the mechanisms of serial recall and the interpretation of the word-length effect follow from the present results. First, these results, along with our previous findings (Hulme et al., 2004), effectively refute current list-based (e.g., Burgess & Hitch, 1999) and item-based (Brown & Hulme, 1995; Neath & Nairne, 1995) accounts of the word-length effect. Second, the large effects of isolation found here were predicted from a model that sees item distinctiveness as a critical determinant of recall (Hulme et al.). In particular, the fact that isolated long words are actually recalled better than background short words in the same list (or indeed short words in pure lists) is a striking finding and provides support for a critical role of item distinctiveness in accounting for serial recall performance (Hulme et al.; see also Cowan et al., 2003).

The results of the present experiments, in line with the findings reported by Hulme et al. (2004), pose serious problems for list-based accounts of the word-length effect and also refute a number

of influential item-based accounts of the effect. We believe that the pattern of data reported here provides further support for a critical role of item distinctiveness at retrieval as proposed by Hulme et al. When applied to the current data, the simulation employing SIMPLE reported by Hulme et al. cannot capture the greater recall advantage shown by isolated long words compared with isolated short words. This counterintuitive finding thus poses critical problems for any current explanation of the word-length effect. It remains to be seen whether a modified version of the distinctiveness framework as embodied in the SIMPLE model (Brown et al., 2002) and used to account for the pattern of data obtained in our earlier experiments (Hulme et al., 2004) will, with *modifications*, be able to account for these results.

¹ We thank Elisabet Service for this suggestion.

References

- Baddeley, A. D. (1986). *Working memory*. London: Oxford University Press.
- Baddeley, A. D., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior*, 14, 575–589.
- Bireta, T. J., Neath, I., & Surprenant, A. M. (in press). The syllable-based word length effect and stimulus set specificity. *Psychonomic Bulletin & Review*.
- Brown, G. D. A., & Hulme, C. (1995). Modeling item length effects in memory span: No rehearsal needed? *Journal of Memory and Language*, 34, 594–621.
- Brown, G. D. A., Neath, I., & Chater, N. (2002). *A ratio model of scale-invariant memory and identification*. Unpublished manuscript.
- Burgess, N., & Hitch, G. (1999). Memory for serial order: A network model of the phonological loop and its timing. *Psychological Review*, 106, 551–581.
- Caplan, D., Rochon, E., & Waters, G. S. (1992). Articulatory and phonological determinants of word-length effects in span tasks. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, 45(A), 177–192.
- Cowan, N., Baddeley, A., Elliott, E. M., & Norris, J. (2003). List composition and the word-length effect in immediate serial recall: A comparison of localist and globalist assumptions. *Psychonomic Bulletin & Review*, 10, 74–79.
- Cowan, N., Day, L., Saults, S., Keller, T. A., Johnson, T., & Flores, L. (1992). The role of verbal output time in the effects of word length on immediate memory. *Journal of Memory and Language*, 31, 1–17.
- Hendry, L., & Tehan, G. (2005). An item/order trade-off explanation of word length and generation effects. *Memory*, 13, 364–371.
- Hulme, C., Maughan, S., & Brown, G. D. A. (1991). Memory for familiar and unfamiliar words: Evidence for a long-term memory contribution to short-term memory span. *Journal of Memory and Language*, 30, 385–401.
- Hulme, C., Surprenant, A. M., Bireta, T. J., Stuart, G., & Neath, I. (2004). Abolishing the word-length effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30, 98–106.
- Kelley, M., & Nairne, J. S. (2001). von Restorff revisited: Isolation, generation, and memory for order. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27, 54–66.
- LaPointe, L. B., & Engle, R. W. (1990). Simple and complex word spans as measures of working memory capacity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16, 1118–1133.
- Loftus, G. R., & Masson, M. E. J. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review*, 1, 476–490.

- Masson, M. E. J., & Loftus, G. R. (2003). Using confidence intervals for graphically based data interpretation. *Canadian Journal of Experimental Psychology*, 57, 203–220.
- 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.
- 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., & Brown, G. D. A. (2006). SIMPLE: Further explorations of a local distinctiveness model of memory. In B. H. Ross (Ed.), *The psychology of learning and motivation*. (pp. 201–243). San Diego, CA: Academic Press.
- Neath, I., & Nairne, J. S. (1995). Word-length effects in immediate memory: Overwriting trace-decay theory. *Psychonomic Bulletin & Review*, 2, 429–441.
- Schmidt, S. R. (1991). Can we have a distinctive theory of memory? *Memory & Cognition*, 19, 523–542.
- Service, E. (1998). The effect of word length on immediate serial recall depends on phonological complexity, not articulatory duration. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, 51A, 283–304.
- von Restorff, H. (1933). Über die wirkung von bereichsbildungen im spurenfeld. [On the effect of sphere formations in the trace field]. *Psychologische Forschung*, 18, 299–342.

Received June 10, 2005

Revision received December 12, 2005

Accepted December 23, 2005 ■