Backward recall and benchmark effects of working memory

Tamra J. Bireta
College of New Jersey, Ewing, New Jersey

Sheena E. Fry, Annie Jalbert, Ian Neath, and Aimée M. Surprenant
Memorial University of Newfoundland, St. John's, Newfoundland, Canada

Gerald Tehan
University of Southern Queensland, Toowoomba, Queensland, Australia

AND

Georgina Anne Tolan
Australian Catholic University, Brisbane, Queensland, Australia

Working memory was designed to explain four benchmark memory effects: the *word length effect*, the *irrelevant speech effect*, the *acoustic confusion effect*, and the *concurrent articulation effect*. However, almost all research thus far has used tests that emphasize forward recall. In four experiments, we examine whether each effect is observable when the items are recalled in reverse order. Subjects did not know which recall direction would be required until the time of test, ensuring that encoding processes would be identical for both recall directions. Contrary to predictions of both the primacy model and the feature model, the benchmark memory effect was either absent or greatly attenuated with backward recall, despite being present with forward recall. Direction of recall had no effect on the more difficult conditions (e.g., long words, similar-sounding items, items presented with irrelevant speech, and items studied with concurrent articulation). Several factors not considered by the primacy and feature models are noted, and a possible explanation within the framework of the SIMPLE model is briefly presented.

Four effects—the *word length effect*, the *irrelevant speech effect*, the *acoustic confusion effect*, and the *concurrent articulation effect*—have played a prominent role in the development of influential theories of immediate memory. Indeed, accounting for these four findings was one of the motivations for creating the phonological loop component of working memory (Baddeley, 1992), and these effects are seen as key data that computational models of short-term memory must account for (Lewandowsky & Farrell, 2008). Despite the numerous studies examining these phenomena, very few studies have examined them using backward recall. To that end, one purpose of the four experiments reported here was to assess whether the four benchmark effects of working memory are observable with backward recall. A second purpose was to test the predictions of two models of memory: Despite their many differences, both the primacy model (Page & Norris, 1998) and the feature model (Nairne, 1990) predict that all four effects should be observed with backward recall.

Empirical Review

Word length effect. The word length effect refers to the finding that lists of short (i.e., one-syllable) words

are recalled better than otherwise comparable lists of longer (i.e., multisyllabic) words (Baddeley, Thomson, & Buchanan, 1975; for a review, see Neath & Surprenant, 2003). The standard paradigm is forward immediate serial recall, but the word length effect is also observable with reconstruction of order (Nairne, Neath, & Serra, 1997), serial recognition (Baddeley, Chincotta, Stafford, & Turk, 2002), free recall (Watkins, 1972), single-item probe recall (Avons, Wright, & Pammer, 1994), and complex span (Tehan, Hendry, & Kocinski, 2001) tests. However, only a small number of studies have examined whether the effect is observable with backward recall.

Cowan et al. (1992, Experiment 3) had subjects recall lists of short and long words in both a forward and a backward order, and recall direction was not known until test. However, a straightforward interpretation of the results is difficult, since word length was manipulated within a list (i.e., the first half of the list was short words, the second half long words), and since the lists had five items, there were not equal numbers of short and long items per list. Moreover, the stimuli used have since been shown to be atypical (see, e.g., Neath, Bireta, & Surprenant, 2003). Cowan, Wood, and Borne (1994) also used mixed lists of

I. Neath, ineath@mun.ca

short and long words, but some lists had just short words and some just long words. An effect of word length was observed with spoken backward recall, but it is difficult to assess the magnitude of the effect, because there was no forward recall condition.

Walker and Hulme (1999) asked subjects to recall lists of short and long words in backward serial order. Auditory presentation was used, spoken recall was required, and the lists contained either concrete or abstract words. There was a clear word length effect for abstract words (0.56 short vs. 0.48 long), but less evidence of such an effect for concrete words (0.63 short vs. 0.58 long). Although a word length effect was demonstrated with backward recall of abstract words, it is unclear whether the effect is of the same magnitude as with forward serial recall, because there was no forward condition.[1]

One can view two experiments, one reported by Cowan, Wood, Nugent, and Treisman (1997) and the other by Cowan, Nugent, Elliott, and Geer (2000), as a between-subjects manipulation of recall direction. Although some lists contained both one- and two-syllable words, as well as words with short or long pronunciation time, there were some lists that contained only one-syllable words, and others that contained only two-syllable words. With spoken forward recall, there was a large effect of number of syllables (see Cowan et al., 2000, Figure 2) but with backward recall, recall of one- and two-syllable words was equivalent (see Cowan et al., 1997, Figure 2, panel C).

Finally, Tehan and Mills (2007) asked subjects to recall four-item lists of short (one-syllable, three phonemes) and long (two- or three-syllable, seven phonemes) words in reverse order. There was no effect of word length, with equivalent recall of short and long items.

Irrelevant speech effect. The irrelevant speech effect (also known as the *irrelevant sound effect* and the *unattended speech effect*) is the finding that recall of visually presented items is disrupted if presentation is accompanied by irrelevant auditory stimuli such as speech (Colle & Welsh, 1976; for a review, see Neath, 2000). As with studies on the word length effect, most researchers use immediate serial recall, but the effect is observed with reconstruction of order tests as well as other tests that involve a serial component (e.g., Farley, Neath, Allbritton, & Surprenant, 2007; Neath, Guérard, Jalbert, Bireta, & Surprenant, 2009). It is clear that this effect is not simply due to distraction, since predictions based on such an account have not fared well empirically (see the discussion in Neath et al., 2009). Despite a large literature on the effect, we could find none that examined the effect of irrelevant speech using backward recall.

Acoustic confusion effect. The acoustic confusion effect (also known as the *phonological similarity effect*) refers to the finding that lists of items that sound similar are recalled worse than lists of otherwise comparable items that sound dissimilar (Conrad, 1964). As with the other phenomena under consideration, it is observable with a variety of different tests, including serial recall, reconstruction of order (Surprenant, Neath, & LeCompte, 1999), free recall (Watkins, Watkins, & Crowder, 1974), and complex span tasks (Tehan et al., 2001). Only two studies examined

backward recall of dissimilar- and similar-sounding items, and they reported contradictory findings.

Rosen and Engle (1997) asked subjects to recall lists of visually presented dissimilar- or similar-sounding letters, in either forward or backward order. The subjects were aware of the recall order prior to presentation of the list. One group of subjects saw only dissimilar-sounding letters and a second saw only similar-sounding letters. The lists varied in length from two to nine letters, and each subject had three lists of each length. Recall was measured in terms of span (i.e., adding up the number of items in a list that were recalled correctly), and responses were made by typing. An acoustic confusion effect was observed in both the forward recall and backward recall groups. Although backward recall reduced span, there was no interaction between similarity and recall direction.

In contrast, Tehan and Mills (2007) used backward recall of words and manipulated the similarity of the words within subjects, but included only backward recall. Recall of dissimilar- and similar-sounding items was equivalent.

Concurrent articulation effect. Concurrent articulation (also known as *articulatory suppression*) is when subjects are asked to say out loud some word, phrase, or stream of items otherwise irrelevant to the task. The result is that overall performance decreases relative to a silent control condition (Murray, 1968), although much of the work has focused on interactions between concurrent articulation and other effects, particularly those noted above (see Neath & Surprenant, 2003, for a review).

Hutton and Towse (2001) measured span for 8- and 11-year-old children, both forward and backward, and both with and without concurrent articulation. The children spoke their responses. Concurrent articulation reduced span scores for forward but not backward recall. Szmalec, Vandierendonck, and Kemps (2005) compared forward and backward recall of letters with and without concurrent articulation, but they report their results in terms of "correct relative order" rather than the more usual strict serial order. They report similar effects of concurrent articulation in both recall directions.

Summary of empirical findings. The few studies listed above differ in numerous ways, including whether memory was assessed by span or by serial recall, whether subjects had foreknowledge of the direction of recall, whether both a forward and backward manipulation were conducted, and whether recall was spoken or written. It is difficult to conclude whether the four benchmark effects of working memory are observable with backward recall and, if they are, whether the effects are smaller or of equivalent size.

Theoretical Review

When Lewandowsky and Farrell (2008, p. 4) examined eight contemporary models of short-term memory, they noted that "somewhat surprisingly, few models (if any) accommodate backward recall." The only model listed that offers an account of both forward and backward recall is the primacy model (Page & Norris, 1998).[2] Although there are numerous models that address a subset of the four benchmark working memory phenomena, there are

only two models that address all four: the primacy model and the feature model (Nairne, 1990).

The primacy model. The primacy model (Page & Norris, 1998, 2003) is a computational model based on the phonological loop of Baddeley (1992). Unlike any other account, the primacy model addresses not only all four benchmark working memory phenomena, but also both forward and backward recall.

The model assumes that the strength of the activation of successive list items decreases over list positions. This results in a primacy gradient, where the activation is the mechanism retaining order information. Recall proceeds by choosing the most active item and then suppressing that item once it has been recalled. This produces the basic serial position function. The word length effect is explained using similar reasoning to that of Baddeley (1992): Within the model, rehearsal serves to offset the decay of the activation; and, because long words take longer to rehearse than do short words, their memory representations are subject to increased decay.

The acoustic confusion effect is the result of a two-stage process. The first stage essentially stores the order of presentation of the items without regard to their similarity. The second stage is an "output stage" (Page & Norris, 1998, p. 772), during which an item selected for output is compared with representations of possible items that do include phonological information. At this point, errors are most likely based on similarity of the phonemes.

The irrelevant speech effect is explained by invoking a second primacy gradient, which is caused by the presence of a second stream of objects (i.e., the items in the irrelevant stream of speech). This second gradient reduces resources available to the first, and is implemented in the model by multiplying the activation levels of the to-be-remembered items by a number less than one. The reduction in activation levels causes a reduction in the proportion of items recalled.

Concurrent articulation reduces recall in the model by preventing rehearsal. As noted above, rehearsal is seen as offsetting the decay, and when this ability to offset decay is removed, overall performance decreases but—because rehearsal is presented equally in all cases—the main effects should still remain.

Within the primacy model, backward serial recall is implemented by assuming that people essentially perform a series of forward recalls; that is, the subject recalls each item in the list, beginning with the first item, until the desired item is reached, in which case it is output as the response. Then, the subject again recalls the list from the first item, but now stops one item earlier and outputs that item. Because the repeated forward recalls act like rehearsal within the model, the primacy gradient is maintained at a high level. This allows the model to predict quite high levels of recall performance in backward recall conditions (see, for example, the simulation reported by Page & Norris, 1998, p. 772). Because backward recall is basically multiple forward recalls that maintain the activation levels of the primacy gradient, the model predicts that all four effects should be observed with backward recall,

and that the magnitude of the effects should be quite comparable, as with forward recall.

Thomas, Milner, and Haberlandt (2003) reported evidence that they took as supporting the idea of multiple covert forward recalls, the process used by the primacy model to account for backward recall. Thomas et al. measured the time to output each response for both forward and backward recall and observed flat response times (RTs) following output of the first item in forward recall, but decreasing RTs for backward recall. They argued that this pattern of RTs is consistent with the idea of a single recall for forward recall and multiple recalls for backward recall. If subjects are recalling a list from first to last (i.e., forward serial recall), they need do so only once, so output times are essentially constant, once retrieval begins. In contrast, if subjects are performing multiple forward recalls (i.e., backward recall), output times should speed up, because the list that the subject needs to traverse gets shorter and shorter.

The feature model. Like the primacy model, the feature model (Nairne, 1990; Neath, 2000; Neath & Nairne, 1995) also offers an account of all four benchmark working memory phenomena. Within the model, items are represented as vectors of features, and memory depends on successfully matching a degraded cue of an item in primary memory with its intact representation in secondary memory. Word length effects are seen as item-level effects: Long words have more segments than short words do. If one assumes a fixed probability of an error in assembling segments, then overall, longer words will be recalled less well than will short words. The acoustic confusion effect occurs because having similar items in a list reduces the probability of successfully matching a degraded cue to the correct intact representation. Both the irrelevant speech and concurrent articulation effects are seen as adding noise, which degrades the cue through a process called *feature adoption*: A feature from the irrelevant item is adopted by the cue, thus reducing the cue's ability to elicit the correct item.

Unlike the primacy model, the feature model does not explicitly provide a description of how backward serial recall would work. Forward serial recall proceeds by using the cue for the first item and determining the best match. There is no reason why recall could not begin with the cue for the last item, however; under this scheme, the same four effects are predicted to be observed, just as with forward recall. The reason is that most of what drives each benchmark working memory effect occurs prior to retrieval, so reversing the order of retrieval will have essentially no effect on the main effect.

Overview of Experiments

The four experiments used a common design, differing only in the key manipulation. Recall direction was manipulated within subjects, and the direction was not revealed until after list presentation. The reason for this type of manipulation is to ensure that whatever processing was performed by the subjects during list presentation, it was likely to be the same for both recall directions. Any differences in processing must necessarily occur after the recall direction is announced. Furthermore, if a benchmark ef-

fect was observed with forward recall, it is evidence that the processing up until the time of the test was sufficient to give rise to the effect. With knowledge of the recall direction prior to encoding, it is possible that subjects may adopt different strategies for forward and backward conditions, so one cannot make the same assumption that the processing at encoding was sufficient to give rise to the effect, should one be observable.

EXPERIMENT 1

Experiment 1 was designed to assess whether a word length effect obtains with both forward and backward recall. Subjects saw seven-item lists of short (one-syllable) or long (three-syllable) words, and after list presentation was over, were informed whether to recall the words in forward or backward order. We used a strict serial reconstruction of order test rather than a written or spoken immediate serial recall test, because it permits output time to be equated: Unlike written or spoken recall, it takes the same amount of time to click on a button labeled with a long word as on a button labeled with a short word (Jalbert, Neath, Bireta, & Surprenant, 2009).

Method

Subjects. Twenty undergraduates from Memorial University of Newfoundland volunteered to participate in exchange for a small honorarium. All identified themselves as native speakers of Canadian English.

Stimuli. The to-be-remembered stimuli were 15 short (one-syllable) and 15 long (three-syllable) words (see the Appendix) equated for frequency, concreteness, imageability, and familiarity, using the MRC psycholinguistic database (Coltheart, 1981). For each list, 7 words were randomly sampled from the appropriate pool and shown in random order.

Design. There were two within-subjects variables, each with two levels: type of list (short or long words) and direction of recall (forward or backward).

Procedure. Subjects were informed that the purpose of the study was to see how accurately they could remember the order in which they saw a list of items. On each trial, each word was displayed for 1 sec in black 28-point Helvetica against a white background on a computer screen. At test, seven buttons appeared on the screen, labeled, in alphabetical order, with the seven words just seen. Simultaneously, the word "Forward" or "Backward" appeared above the response buttons. The subjects were asked to click on the buttons to re-create either the original presentation order if they saw the word "Forward" or the reverse order if they saw the word "Backward." That is, with forward recall, they clicked on the first word first, the second word second, and so on. With backward recall, they clicked on the last word first, the penultimate word second, and so on. Once a response had been made, it could not be changed.

There were 40 experimental trials, 10 in each condition. Assignment of trials to condition was randomized for each subject. Each subject was tested individually and the experimenter remained in the room to ensure that the instructions were followed.

Results

The main results, displayed in the upper left panel of Figure 1, illustrate that a word length effect was observed for forward but not for backward recall.[3] With backward
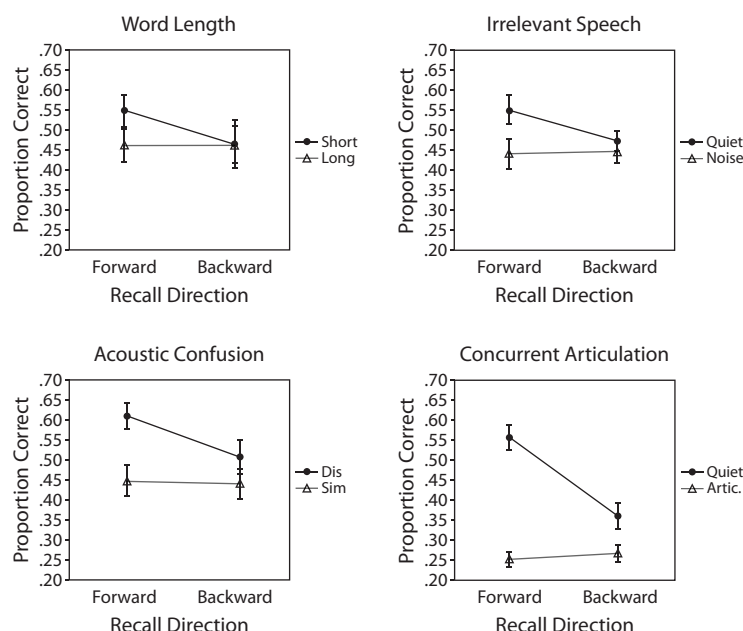
Figure 1. Proportion of short and long words recalled as a function of recall direction (upper left); proportion of letters recalled as a function of presence or absence of irrelevant speech and recall direction (upper right); proportion of letters recalled as a function of acoustic confusability and recall direction (lower left); and proportion of words recalled as a function of presence or absence of concurrent articulation and recall direction (lower right). Error bars show the standard errors of the means. Dis, dissimilar; Sim, similar; Artic., concurrent articulation.
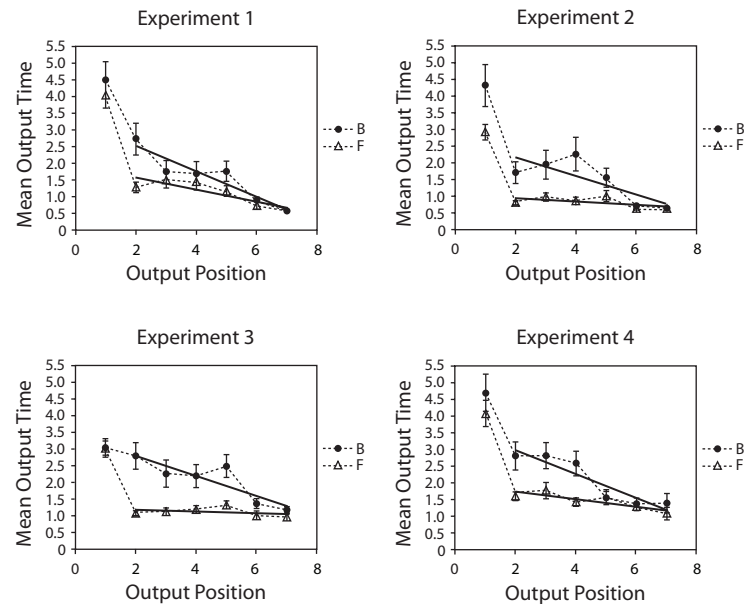
Figure 2. Mean output time (in seconds) for items correctly recalled as a function of recall direction and output serial position in Experiment 1 (upper left), Experiment 2 (upper right), Experiment 3 (lower left), and Experiment 4 (lower right). Error bars show the standard errors of the means. B, backward recall; F, forward recall.

recall, the proportion of short and long items recalled was equivalent. A 2 (word length: short or long) × 2 (recall direction: forward or backward) × 7 (serial position) repeated measures ANOVA was conducted, and it supported these observations.

There was a significant main effect of length [$F(1,19) = 5.412$, $MS_e = 5.074$, $p < .05$], with better recall of short than of long words (0.507 vs. 0.463). The main effect of recall direction was not significant [$F(1,19) = 1.972$, $MS_e = 11.77$, $p > .15$], although there was a trend toward better recall with forward than with backward instructions (0.505 vs. 0.464). There was also a significant effect of serial position [$F(6,114) = 12.803$, $MS_e = 3.343$, $p < .001$].

Critically, there was an interaction between length and recall direction [$F(1,19) = 9.162$, $MS_e = 2.623$, $p < .01$], with a word length effect apparent with forward recall (0.548 vs. 0.466) but no effect with backward recall (0.462 vs. 0.463). Tukey HSD tests revealed that performance in the short-forward condition was significantly better than in the three other conditions, which did not differ; that is, there was no word length effect with backward recall, and backward recall had no effect on recall of long items.

The only other significant interaction was between recall direction and position [$F(6,114) = 22.354$, $MS_e = 24.05$, $p < .001$]. This was due to superior recall of early items compared with late items with forward recall, and the reverse with backward recall. The interaction between length and position and the three-way interaction were not significant ($Fs < 1$).

During recall, the computer recorded how long the subject took to make each response. Figure 2 shows the mean output time for items correctly recalled at each position as

a function of recall direction. As can be seen in the upper left panel of Figure 2, backward recall took slightly longer than did forward recall. The best-fitting line for Positions 2–7 is also shown. Both forward and backward recall exhibit negative slopes, although the slope is steeper for backward recall (see Table 1). A 2 (word length: short or long) × 2 (recall direction: forward or backward) × 7 (serial position) repeated measures ANOVA was conducted on the output time data.[4]

There was no difference in output times as a function of word length ($F < 1$), with almost identical mean output times for short and long items (1.785 vs. 1.754 sec, respectively). There was a significant main effect of direction [$F(1,19) = 16.482$, $MS_e = 1.726$, $p < .001$], with faster output for forward than for backward recall (1.544 vs. 1.995 sec, respectively). There was also a significant main effect of position [$F(6,114) = 69.416$, $MS_e = 1.689$, $p < .001$], indicating faster responding for later than for earlier items. All interactions had $F < 1$, except for the interaction

Table 1
Slopes of Best-Fitting Lines for Output Positions 2–7 for Forward and Backward Recall, and Measures of Fit (R^2 and Root Mean Square Deviation [RMSD]) for Experiments 1–4

	Experiment			
	1	2	3	4
Forward	−0.175	−0.051	−0.023	−0.110
R^2	0.732	0.333	0.112	0.689
RMSD	0.170	0.108	0.095	0.112
Backward	−0.380	−0.280	−0.297	−0.352
R^2	0.874	0.623	0.768	0.856
RMSD	0.200	0.313	0.195	0.235

between direction and position [$F(6,114) = 2.436$, $MS_e = 2.012$, $p < .05$]. This was due mostly to slower output times initially for backward than for forward recall, but this difference disappeared for the last few items output.

Discussion

When word length and recall direction were manipulated within subjects, and when those subjects did not know the recall direction during list presentation, a standard word length effect was seen in forward recall but was absent in backward recall. This lack of a word length effect with backward recall replicates Tehan and Mills (2007). Both the primacy model and the feature model predict that a word length effect should have been seen with both recall directions. As is apparent in Figure 2, backward recall reduced recall of short items but had no effect on recall of long items.

Backward recall took longer than forward recall, but there was no difference in output times as a function of word length; this latter finding replicates Jalbert et al. (2009). The pattern of results differs from that reported by Thomas et al. (2003) in two respects. First, forward recall RTs for Output Positions 2–7 decreased (slope = −0.175), whereas those of Thomas et al. were flat. Second, backward recall RTs showed a largely linear trend (although the slope, −0.380, was approximately twice as steep as forward recall), whereas Thomas et al. observed an inverted U-shaped function. Rather, the results for backward recall are more similar to those reported by Anderson, Bothell, Lebiere, and Matessa (1998), in that the time to output the first response is longer than the time to output the succeeding responses. One possible reason for the difference is the method of responding. The subjects in the Thomas et al. study had to type in their responses (nouns up to seven letters long), whereas the subjects in the present study clicked on buttons. A full consideration of output times is postponed until the General Discussion.

EXPERIMENT 2

Experiment 2 was the same as Experiment 1, except that instead of manipulating word length, we manipulated whether presentation of the to-be-remembered items was accompanied by irrelevant background speech. Subjects saw seven-item lists of letters presented in quiet or with irrelevant speech, and after list presentation was over, were signaled whether to recall the letters in forward or backward order.

Method

Subjects. Twenty undergraduates from Memorial University of Newfoundland volunteered to participate in exchange for a small honorarium. None had participated in Experiment 1, and all identified themselves as native speakers of Canadian English.

Stimuli. The to-be-remembered stimuli were eight letters (F, K, L, M, Q, R, and S) used by Colle and Welsh (1976) in their study of the irrelevant speech effect. The irrelevant speech was a passage in German from *Die Wilden* by Franz Kafka spoken by a female; this passage has previously been used to produce reliable irrelevant speech effects (Neath, Farley, & Surprenant, 2003).

Design. There were two within-subjects variables, each with two levels: presence of irrelevant speech (present or absent) and direction of recall (forward or backward).

Procedure. The procedure was the same as that in Experiment 1, except for the stimuli and the manipulation of irrelevant speech, which occurred on half the trials. The irrelevant speech was played over headphones and began at the onset of the trial, and ended with the offset of the final to-be-remembered item.

Results

As can be seen in the upper right panel of Figure 1, an irrelevant speech effect was observed for forward recall, but with backward recall there was no detrimental effect of irrelevant speech. A 2 (noise condition: quiet or irrelevant speech) × 2 (recall direction: forward or backward) × 7 (serial position) repeated measures ANOVA was conducted and supported these observations.

There was a significant main effect of noise [$F(1,19) = 21.610$, $MS_e = 2.983$, $p < .001$], with better recall in the quiet than in the irrelevant speech condition (0.512 vs. 0.444). The main effect of recall direction was not significant [$F(1,19) = 1.932$, $MS_e = 9.616$, $p > .15$], although there was a trend toward better recall with forward rather than with backward instructions (0.496 vs. 0.460). There was also a significant effect of serial position [$F(6,114) = 21.684$, $MS_e = 2.816$, $p < .001$].

Critically, there was an interaction between noise and recall direction [$F(1,19) = 4.903$, $MS_e = 4.901$, $p < .05$], with an irrelevant speech effect apparent with forward recall (0.551 vs. 0.441) but no effect with backward recall (0.472 vs. 0.464). Tukey HSD tests revealed that performance in the quiet-forward condition was significantly better than in the three other conditions, which did not differ; that is, there was no irrelevant speech effect with backward recall, and backward recall had no effect on recall of items when presentation was accompanied by irrelevant speech.

The only other significant interaction was between recall direction and position [$F(6,114) = 22.352$, $MS_e = 2.159$, $p < .001$]. This was due to superior recall of early items compared with late items with forward recall, and the reverse with backward recall. The interaction between noise and position was not significant [$F(6,114) = 1.726$, $MS_e = 1.993$, $p > .10$], and neither was the three-way interaction ($F < 1$).

As can be seen in the upper right panel of Figure 2, backward recall took longer than forward recall. Unlike in Experiment 1, there was no discernible slope for Positions 2–7 with forward recall, and backward recall exhibited more of an inverted U-shaped function. A 2 (noise condition: quiet or irrelevant speech) × 2 (recall direction: forward or backward) × 7 (serial position) repeated measures ANOVA was conducted on the output time data.

There was no difference in output times as a function of noise condition ($F < 1$), with overall output times equivalent for quiet and irrelevant speech conditions (1.517 vs. 1.523 sec, respectively). There was a significant main effect of direction [$F(1,19) = 24.953$, $MS_e = 3.152$, $p < .001$], with faster output for forward than for backward recall (1.145 vs. 1.894 sec, respectively). There was also a significant main effect of position [$F(6,114) = 43.015$,

$MS_e = 1.867$, $p < .001$]. As in Experiment 1, the only significant interaction was between direction and position [$F(6,114) = 3.193$, $MS_e = 2.022$, $p < .01$]. This was due mostly to slower output times initially for backward as compared with forward recall, but this difference disappeared for the last few items output. All other interactions had $F < 1$, except for the three-way interaction [$F(6,114) = 1.018$, $MS_e = 0.839$, $p > .40$].

Discussion

When the presence or absence of irrelevant speech during list presentation and recall direction were manipulated within subjects, and when those subjects did not know recall direction during list presentation, a standard irrelevant speech effect was seen in forward recall, but was absent in backward recall.

The pattern of output times differed from those observed in Experiment 1; instead, they were more reminiscent of those reported by Thomas et al. (2003). In particular, forward recall was almost flat (slope = −0.051), and backward recall exhibited an inverted U shape.

EXPERIMENT 3

Experiment 3 was essentially the same as Experiments 1 and 2, except that instead of manipulating word length or irrelevant speech, we manipulated whether the to-be-remembered items were similar- or dissimilar-sounding letters. Subjects saw seven-item lists of similar- or dissimilar-sounding letters and, after list presentation was over, were informed whether to recall the letters in forward or backward order.

Method

Subjects. Twenty undergraduates from the College of New Jersey volunteered to participate in exchange for course credit. All identified themselves as native speakers of American English.

Stimuli. The to-be-remembered stimuli were the letters FKLMQRX and BCDGPTV, which have previously been used to produce a robust acoustic confusion effect (e.g., Surprenant et al., 1999).

Design. There were two within-subjects variables, each with two levels: similarity (dissimilar- or similar-sounding letters) and direction of recall (forward or backward).

Procedure. The procedure was the same as in Experiments 1 and 2, except that the stimuli were dissimilar- and similar-sounding letters.

Results

As can be seen in the lower left panel of Figure 1, an acoustic confusion effect was observed for forward, but not backward, recall. A 2 (similarity: dissimilar- or similar-sounding letters) × 2 (recall direction: forward or backward) × 7 (serial position) ANOVA was conducted and supported these observations.

There was a significant main effect of similarity [$F(1,19) = 21.955$, $MS_e = 0.108$, $p < .001$], with better recall in the dissimilar than in the similar condition (0.558 vs. 0.444). Unlike in Experiments 1 and 2, the main effect of recall direction was significant [$F(1,19) = 4.519$, $MS_e = 9.009$, $p < .05$], with better recall following forward rather than backward instructions (0.528 vs.

0.474). There was also a significant effect of serial position [$F(6,114) = 13.765$, $MS_e = 3.067$, $p < .001$].

Critically, there was an interaction between similarity and recall direction [$F(1,19) = 6.421$, $MS_e = 4.772$, $p < .05$], with a substantial effect of similarity with forward recall (0.609 vs. 0.447) and a greatly reduced effect with backward recall (0.508 vs. 0.440). Tukey HSD tests revealed that performance in the dissimilar-forward condition was significantly better than in the three other conditions, which did not differ; that is, recall in the dissimilar condition was better with forward than with backward instructions, but recall in the similar condition was unaffected by recall direction.

There was a significant interaction between recall direction and position [$F(6,114) = 78.807$, $MS_e = 2.420$, $p < .001$]. This was due to better recall of early items than of late items with forward recall, and the reverse with backward recall. The interaction between similarity and position was significant [$F(6,114) = 3.022$, $MS_e = 1.342$, $p < .01$]. The three-way interaction was not reliable [$F(6,114) = 1.277$, $MS_e = 1.487$, $p > .25$].

As can be seen in the lower left panel of Figure 2, backward recall took longer than forward recall did, and more closely followed the pattern observed in Experiment 2 than that in Experiment 1. A 2 (similarity: dissimilar- or similar-sounding letters) × 2 (recall direction: forward or backward) × 7 (serial position) repeated measures ANOVA was conducted on the output time data.

There was no difference in output times for dissimilar and similar letters ($F < 1$), with overall output times equivalent for quiet and irrelevant speech conditions (1.881 vs. 1.957 sec, respectively). There was a significant main effect of direction [$F(1,19) = 47.213$, $MS_e = 1.438$, $p < .001$], with faster output for forward than backward recall (1.487 vs. 2.351 sec, respectively). There was also a significant main effect of position [$F(6,114) = 19.610$, $MS_e = 1.284$, $p < .001$]. As in Experiments 1 and 2, the only significant interaction was between direction and position [$F(6,114) = 6.178$, $MS_e = 1.244$, $p < .01$]. All other interactions were $F < 1$, except for similarity and position [$F(6,114) = 1.129$, $MS_e = 0.965$, $p > .35$].

Discussion

When acoustic similarity and recall direction were manipulated within subjects, and when those subjects did not know recall direction during list presentation, a standard acoustic confusion effect was seen in forward recall, but was absent in backward recall. These results replicate Tehan and Mills (2007), but differ from those reported by Rosen and Engle (1997). The main differences between the latter study and both the current experiment and that reported by Tehan and Mills is that Rosen and Engle used span as a measure rather than proportion correct at each serial position, and also manipulated acoustic similarity between subjects.

The output times resemble those seen in Experiment 2 and in Thomas et al. (2003), with essentially flat responding over Positions 2–7 with forward recall (slope = −0.023) and an inverted U-shaped function with backward recall.

EXPERIMENT 4

Experiment 4 was essentially the same as the three previous experiments, except that we manipulated whether the subjects engaged in concurrent articulation during presentation of the to-be-remembered items.

Method

Subjects. Twenty undergraduates from the College of New Jersey volunteered to participate in exchange for course credit. None had participated in Experiment 3, and all identified themselves as native speakers of American English.

Stimuli. The to-be-remembered stimuli were the words *aisle, arch, branch, club, dance, edge, flute, globe, inch, learn, ounce, priest, quick, ridge, scotch, smile, snail, split, sponge, trial, troop, view, voice,* and *wreck*. For the concurrent articulation manipulation, the subjects were asked to say the numbers 1, 2, 3, and 4 out loud, over and over during list presentation.

Design. There were two within-subjects variables, each with two levels: presence of concurrent articulation (present or absent) and direction of recall (forward or backward).

Procedure. The procedure was the same as in the three previous experiments, except that the stimuli were different words and the subjects were asked to engage in concurrent articulation on half the trials. Immediately prior to each trial, a message indicated whether the subject should engage in concurrent articulation, and after the final item had been shown, a message informed the subjects that they could end articulation for that trial.

Results

As can be seen in the lower right panel of Figure 1, a concurrent articulation effect was observed for both forward and backward recall, although the effect was attenuated with backward recall instructions. A 2 (articulation condition: quiet or concurrent articulation) × 2 (recall direction: forward or backward) × 7 (serial position) repeated measures ANOVA was conducted.

There was a significant main effect of articulation [$F(1,19) = 71.250$, $MS_e = 7.692$, $p < .001$], with better recall in the quiet than in the concurrent articulation condition (0.458 vs. 0.260). There was also a main effect of recall direction [$F(1,19) = 14.934$, $MS_e = 7.715$, $p < .001$], with better recall with forward rather than with backward instructions (0.404 vs. 0.312). There was also a significant effect of serial position [$F(6,114) = 19.298$, $MS_e = 3.155$, $p < .001$].

Critically, there was an interaction between articulation and recall direction [$F(1,19) = 44.622$, $MS_e = 3.459$, $p < .001$], with a substantial effect of concurrent articulation with forward recall (0.556 vs. 0.253) and a reduced effect with backward recall (0.360 vs. 0.267). Tukey HSD tests revealed that performance in the quiet-forward condition was significantly better than in the three other conditions. In addition, recall in the quiet-backward condition was better than in both the concurrent articulation-forward and the concurrent articulation-backward conditions, which did not differ; that is, although the benchmark effect of concurrent articulation was observed with both forward and backward recall, it was smaller in the backward recall condition, and recall in the concurrent articulation condition was unaffected by recall direction.

The only other significant interaction was between recall direction and position [$F(6,114) = 35.610$, $MS_e = 2.064$, $p < .001$]. This was due to superior recall of early items compared with late items with forward recall, and the reverse with backward recall. The interaction between articulation and position was not significant [$F(6,114) = 1.627$, $MS_e = 1.679$, $p > .10$]. The three-way interaction was reliable [$F(6,114) = 10.921$, $MS_e = 1.682$, $p < .001$].

As can be seen in the lower right panel of Figure 2, backward recall took longer than did forward recall, and the pattern was more similar to that seen in Experiment 1 than to those seen in Experiments 2 and 3. Both forward and backward recall showed a negative slope, and the pattern in the backward condition did not exhibit an inverted U shape. A 2 (articulation condition: quiet or concurrent articulation) × 2 (recall direction: forward or backward) × 7 (serial position) repeated measures ANOVA was conducted on the output time data.

There was no difference in output times as a function of whether the subject engaged in concurrent articulation ($F < 1$), with overall output times equivalent for quiet and articulation conditions (1.785 vs. 1.755 sec, respectively). There was a significant main effect of direction [$F(1,19) = 16.482$, $MS_e = 1.726$, $p < .001$], with faster output for forward than for backward recall (1.544 vs. 1.995 sec, respectively). There was also a significant main effect of position [$F(6,114) = 69.416$, $MS_e = 1.689$, $p < .001$]. As in Experiments 1–3, the only significant interaction was between direction and position [$F(6,114) = 2.436$, $MS_e = 2.012$, $p < .05$]. All other interactions were $F < 1$.

Discussion

Unlike the three previous experiments, Experiment 4 found an effect of a benchmark manipulation on backward recall, although the detrimental effect of concurrent articulation was smaller for backward than for forward recall. Similar to the previous experiments, though, backward recall had no effect on the more difficult condition; that is, performance under concurrent articulation was equivalent regardless of recall direction.

The output time data were similar to those in Experiment 1, and different from those in Experiments 2 and 3. Both showed negative slopes from Positions 2 to 7, with a steeper slope for backward than for forward recall (-0.352 vs. -0.110, respectively). One obvious difference is that Experiments 1 and 4 used words, whereas Experiments 2 and 3 used letters. However, Thomas et al. (2003) also used words, but their data were more similar to those seen here with letters.

GENERAL DISCUSSION

In the four experiments, subjects were unaware of the recall direction—forward or backward—until after all of the to-be-remembered items had been presented. When the signal was to recall in strict forward order, all four working memory benchmark effects were observed. However, a very different pattern was observed when the signal was to recall in strict backward order. With backward recall, there was no word length effect and no irrelevant speech effect, the acoustic confusion effect failed to reach conventional levels of significance, and the concurrent ar-

ticulation effect, although present, was greatly attenuated. Backward recall, then, either eliminates or attenuates all four benchmark working memory phenomena.

A similar effect of recall direction was observed by Hulme et al. (1997). They tested forward and backward serial recall of seven-item lists of high- and low-frequency words. With forward recall, a standard word frequency effect was observed. With backward recall, however, the effect of word frequency was absent. Moreover, most of the change with backward recall was a decrease in recall of the high-frequency words. As with the present experiments, recall of the more difficult condition was roughly comparable, whether the recall direction was forward or backward.

The output time data are less orderly. Thomas et al. (2003) reported flat output times after the first item, but we observed negative slopes in all experiments. Consistent with the findings of Thomas et al., we observed steeper slopes for backward recall than for forward recall. However, this does not necessarily provide evidence in support of the multiple covert forward recall view, since two different patterns emerged (Experiments 1 and 4, and Experiments 2 and 3).

Explaining Backward Recall

The results are clearly at odds with the predictions of the primacy model and the feature model. Both models predicted that the benchmark effects should be observed with backward recall. The only effect that was preserved with backward recall was the detrimental effect of concurrent articulation, but even this effect was greatly attenuated.

Primacy model. One way that the primacy model could be altered to account for the effects observed (or more accurately, the effects not observed) is to assume that subjects employ different strategies depending on the recall direction. For example, the requirement to recall in backward order may lead to the abandonment of the phonological loop. If so, one would expect effects that rely on phonological processing to disappear. There is some evidence that suggests that people may use different processing in the two directions. For example, one account of backward recall, called the *representation view*, suggests that phonological representations are well suited for retaining forward order, whereas a visuospatial representation might be more appropriate for reading the items in a backward order (Li & Lewandowsky, 1995). One line of data supporting this view is that tasks designed to interfere with interitem associations used between phonological representations disrupted forward recall but not backward recall.

Nonetheless, incorporating this assumption leads to two problems. First, such an appeal to altered processing strategies was not needed when the primacy model was fit to the forward and backward data of Hulme et al. (1997), even though the subjects in the Hulme et al. (1997) study knew the recall direction prior to list presentation. It is not clear why some tasks lead to altered processing strategies and others do not. Second, an appeal to differing processing strategies seems less plausible when one considers that the subjects in the present experiments did not know the recall direction until the time of recall.

Feature model. Although the feature model has not yet been applied to backward recall, the standard version can be modified so that, rather than initiating recall with the cue for the first item, recall is initiated with the cue for the last item. This results, in the model, in equivalent effects with forward and backward recall.

One way in which the feature model may be able to offer an account of backward recall is to incorporate what Rosen and Engle (1997) termed a *complexity view* of backward recall. According to this view, forward and backward recall rely on the same processing and representation, but backward recall requires an attention-demanding transformation of the stimuli before output. A similar idea is that, whereas forward serial recall is a highly practiced and most likely automatic skill, backward serial recall is something rarely done in everyday life and is a more controlled process. It is possible that such a transformation can be more disruptive to easier tasks than to more difficult tasks. Although this may sound counterintuitive at first, consider the following.

From examining Figure 1, it is apparent that backward recall affects the "easier" conditions (i.e., short words, no irrelevant speech, dissimilar-sounding items, no concurrent articulation) but has little or no effect on the "harder" conditions (i.e., long words, irrelevant speech, similar-sounding items, concurrent articulation). In the feature model, items are represented as vectors of features, some of which can be interfered with via overwriting or feature adoption. With more overwriting or more feature adoption, performance decreases (on average), but only to a certain extent. Some manipulations are conceived of as affecting the same feature location multiple times. For example, an item similar to others will share lots of the same feature values. Because the values are the same, they are not helpful in matching the degraded cue to the correct intact item. If some of these essentially useless items are subject to feature adoption, performance is not necessarily worse because a "useless" feature has been altered to a mismatching feature.

This notion of complexity affecting easier tasks more than harder tasks may explain the apparent contradiction between our results and those of Farrand and Jones (1996), who found that with serial reconstruction of order, there were no differences in level of recall as a function of direction of recall. One difference between the studies is that there was a 10-sec retention interval prior to retrieval in the experiments of Farrand and Jones. A second difference is that our stimuli were always arranged in alphabetical order to facilitate selection, whereas Farrand and Jones used a different random order on each trial. Both differences could make the Farrand and Jones task more difficult than ours. On the theory that backward recall affects easier tasks more than it affects harder tasks, one way of reconciling the difference is that Farrand and Jones's experiment with delayed serial recall was harder than ours, and thus no difference was observed (see also Neale & Tehan, 2007).

TODAM. One model not yet discussed is TODAM (theory of distributed associative memory; Lewandowsky & Murdock, 1989). Although TODAM does not address

the benchmark working memory effects, it does make an interesting general prediction about forward and backward recall. Li and Lewandowsky (1993) showed that TODAM predicts, in general, that variables affecting forward recall should have little or no effect on backward recall. First, Li and Lewandowsky (1993) noted that, in addition to other factors, forward and backward recall differ in terms of output interference. With forward recall, the sequence A B C D E F must be recalled as A B C D E F; that is, five encoding and retrieval events intervene between presentation of an item and recall of that item. With backward recall, each item is followed by a different number of events, from 0 (for F) to 10 (for A). This is incorporated into the model so that, essentially, with backward recall the growing output interference is largely responsible for recall performance (for further details and simulations, see Li & Lewandowsky, 1993). Although it is not clear that this could result in precisely the pattern observed here (i.e., not only no effect of a variable with backward recall, but no overall detriment), it does seem plausible that differential interference could very well be an important factor.

SIMPLE. A final model that might account for the results reported here is SIMPLE (scale independent memory and perceptual learning; Brown, Neath, & Chater, 2007). SIMPLE sees recall as solving a discrimination problem within a multidimensional psychological space. Items within a sparsely populated region of this space are easily discriminated and thus readily recalled, but those items in densely populated regions are harder to discriminate among and are thus less likely to be recalled. The primary dimension in the model involves the temporal characteristics of the task determined by presentation rate and retention interval timing parameters, and it is this dimension that maintains order information.[5] Using a telephone pole analogy, temporal memories become less discriminable as they recede in time, just as telephone poles become less discriminable as they recede into the distance. In other words, the temporal dimension has a profound recency bias; recently presented items tested immediately are highly discriminable, whereas items from earlier in the list are less so.

Many memory effects modeled by SIMPLE are based on a two-dimensional space, with the second dimension depending on the specifics of the experiment. Frequently, the second dimension captures variation among the items. Hendry and Tehan (2005; see also Tehan & Tolan, 2007) have suggested an item/order trade-off account of the word length effect that may well apply to the other benchmark data. The basic thrust of this argument is that items that can be encoded easily at the item level (e.g., short words, dissimilar-sounding items, words without background speech or no concurrent articulation) will benefit with regard to encoding of order. For example, short items can be identified faster than long items can, so there is therefore more time (or more resources) available to process order information, which will lead to better performance on tasks that require memory for order. If the encoding of item information is made more difficult by irrelevant background speech or concurrent articulation, the order encoding advantage is going to be attenuated. Such an explanation may account for the selective effects

on "easy" conditions, but at heart it is an encoding argument. Consequently, in the present experiments, one might have differential encoding of order information in the "easy" conditions, but it seems unclear why the beneficial order encoding does not facilitate both forward and backward recall. Thus, the item/order trade-off account, although potentially explaining differences in "easy" and "difficult" memory effects, fares no better than do other models in accounting for the attenuation of the benchmark effects with backward recall.

The basic idea, however, may be further explicated by examining a particular instantiation. The trade-off between item and order information can be implemented in SIMPLE by altering the attentional weights on the time and item dimensions as a function of experimental task (see also Lewandowsky, Brown, & Thomas, 2009; Lewandowsky, Nimmo, & Brown, 2008). Because the sum of the weights is constrained to be 1.0, if more attention is paid to the temporal dimension, less attention is paid to an item dimension, and vice versa. In short, an a priori prediction of SIMPLE is that an increase in attention to a temporal dimension will result in attenuated item effects in a two-dimensional model. The post hoc explanation for the present results involves the assumption that, with backward recall, participants do vary the degree to which they attend to the temporal dimension at the time of retrieval.

Hulme, Surprenant, Bireta, Stuart, and Neath (2004; see also Hulme et al., 2006) first described how SIMPLE could account for word length effects, but their model required three dimensions. Neath and Brown (2006) demonstrated a revised model that required only two dimensions. The simulation below is based on the latter version.

In SIMPLE, the words are represented on both a temporal and an item dimension. For the latter, the seven short items have values of 1.0, 1.5, 2.0, and so on, whereas the long items have values of 0.1, 0.2, 0.3, and so on. (The order of these values is randomized on each trial of the simulation, and the simulation involved 10,000 trials.) The values on the temporal dimension are determined by the experiment. Each item appeared for 1 sec, so the inter-item presentation interval (IPI) is 1 sec. With forward recall, the first item was recalled, on average, 4 sec after the end of the list (see Figure 2), but for backward recall, this value was 4.5 sec. Similarly, each subsequent item took 1.5 sec for forward and 1.9 sec for backward recall.

Thus, for forward recall, the temporal dimension for Items 1–7 have values of 10, 9, 8, 7, 6, 5, and 4 (respectively) at the time Item 1 is recalled. This is made up of the 4-sec time for the first item to be recalled on average (determined empirically), and the 1-sec IPI separating each item. When an additional item is recalled, these values are incremented by 1.5. At the time Item 7 is recalled, the values are 19, 18, 17, 16, 15, 14, and 13. With backward recall, the same dimension is used, but the temporal dynamics differ (see Brown & Lamberts, 2003). Thus, with backward recall, the temporal dimension for Items 1–7 has the values 10.5, 9.5, 8.5, 7.5, 6.5, 5.5, and 4.5. This is made up of the 4.5-sec time for the first response to be made on average (determined empirically), and the 1-sec IPI separating each
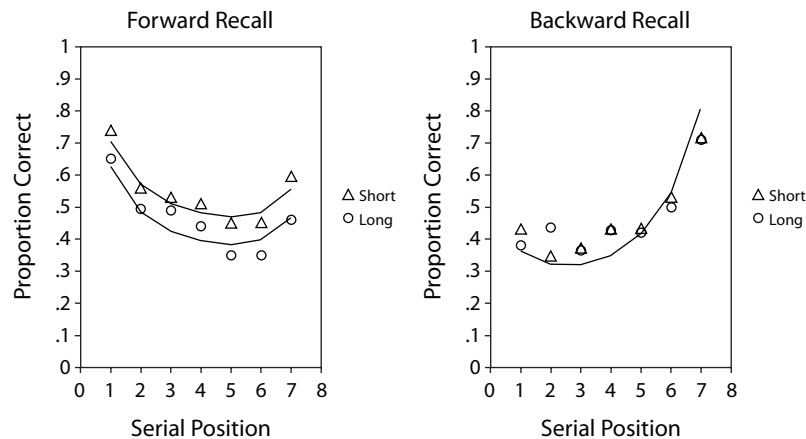
Figure 3. Proportion of short and long words recalled in forward (left panel) or backward (right panel) order (data points) and predictions of SIMPLE (lines). See the text for details.

item. When an additional item is recalled, these values are incremented by 1.9. At the time Item 1 is recalled, the values are 21.9, 20.9, 19.9, 18.9, 17.9, 16.9, and 15.9.

As with many other models, it is assumed that the temporal values undergo a log transformation. The similarity, $\eta_{i,j}$, between two log-transformed memory representations of the relative time of presentation, T_i and T_j, and two representations of item information, I_i and I_j, is given by Equation 1.

$$\eta_{i,j} = e^{-c(W_T|T_i - T_j| + W_I|I_i - I_j|)} \tag{1}$$

The weighting parameters, W_T and W_I, are constrained, such that $W_T + W_I = 1.0$. The two free parameters, then, are c and W_T, although it is acknowledged that the values on the item dimension are arbitrary.

The probability of producing the response associated with item i, R_i, when given the cue for stimulus j, C_j, is given by Equation 2 (i.e., the Luce choice rule), in which n is the number of items in the response set:

$$P\left(R_i \mid C_j\right) = \frac{\eta_{i,j}}{\sum_{k=1}^{n} \eta_{i,k}}. \tag{2}$$

With c set to 8.6 for both forward and backward recall, and W_T set to 0.959 for forward recall and 1.0 for backward recall, SIMPLE produces the results shown in Figure 3 (symbols show the data, lines show the model's predictions).[6] Although not capturing all aspects of the data, the model is reproducing the general pattern quite accurately by varying only the relative weights on the temporal and item dimensions (R^2 was 0.907 for forward and 0.949 for backward). In particular, the shape of the curves, the level of performance, and the presence of the word length effect in only forward recall are all reproduced.

SIMPLE has yet to be applied to the effects of irrelevant speech and concurrent articulation (although see Neath & Brown, 2006). However, assuming that the base of the effect is captured by the item dimension, a similar modeling approach should produce appropriate results.

SUMMARY

Forward recall is characterized by robust influences of word length, acoustic confusability, irrelevant background speech, and concurrent articulation, the so-called *benchmark working memory effects*; backward recall is not subject to these influences to the same extent and in some cases in our data, not at all. Whereas some models can address some aspects of backward recall, no model yet offers a comprehensive account of backward recall that includes an explanation of the interactions observed in this article. At the very least, with the possible exception of SIMPLE, our present theories and models of forward recall seem unlikely to apply to backward recall in any straightforward manner. Moreover, in the wider working memory and neuropsychological literature, backward recall is often depicted as a hybrid of simple span and complex span tasks. This conclusion also seems suspect in that complex span tasks (e.g., operation span and reading span) are susceptible to word length and acoustic confusability effects (Tehan et al., 2001), whereas backward recall is not. The present results suggest that backward recall is a different entity from the class of forward recall working memory tasks.

REFERENCES

ANDERSON, J. R., BOTHELL, D., LEBIERE, C., & MATESSA, M. (1998). An integrated theory of list memory. *Journal of Memory & Language*, **38**, 341-380.

AVONS, S. E., WRIGHT, K. L., & PAMMER, K. (1994). The word-length effect in probed and serial recall. *Quarterly Journal of Experimental Psychology*, **47A**, 207-231.

BADDELEY, A. [D.] (1992). Is working memory working? The fifteenth

Bartlett lecture. *Quarterly Journal of Experimental Psychology*, **44A**, 1-31.

BADDELEY, A. [D.], CHINCOTTA, D., STAFFORD, L., & TURK, D. (2002). Is the word length effect in STM entirely attributable to output delay? Evidence from serial recognition. *Quarterly Journal of Experimental Psychology*, **55A**, 353-369.

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.

BROWN, G. D. A., & LAMBERTS, K. (2003). Double dissociations, models and serial position curves. *Cortex*, **39**, 148-152.

BROWN, G. D. A., NEATH, I., & CHATER, N. (2007). A temporal ratio model of memory. *Psychological Review*, **114**, 539-576.

BROWN, G. D. A., PREECE, T., & HULME, C. (2000). Oscillator-based memory for serial order. *Psychological Review*, **107**, 127-181.

COLLE, H. A., & WELSH, A. (1976). Acoustic masking in primary memory. *Journal of Verbal Learning & Verbal Behavior*, **15**, 17-31.

COLTHEART, M. (1981). The MRC psycholinguistic database. *Quarterly Journal of Experimental Psychology*, **33A**, 497-505.

CONRAD, R. (1964). Acoustic confusions in immediate memory. *British Journal of Psychology*, **55**, 75-84.

COWAN, N., DAY, L. J., SAULTS, J. 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 & Language*, **31**, 1-17.

COWAN, N., NUGENT, L. D., ELLIOTT, E. M., & GEER, T. (2000). Is there a temporal basis of the word length effect? A response to Service (1998). *Quarterly Journal of Experimental Psychology*, **53A**, 647-660.

COWAN, N., WOOD, N. L., & BORNE, D. N. (1994). Reconfirmation of the short-term storage concept. *Psychological Science*, **5**, 103-106.

COWAN, N., WOOD, N. L., NUGENT, L. D., & TREISMAN, M. (1997). There are two word-length effects in verbal short-term memory: Opposed effects of duration and complexity. *Psychological Science*, **8**, 290-295.

FARLEY, L. A., NEATH, I., ALLBRITTON, D. W., & SURPRENANT, A. M. (2007). Irrelevant speech effects and sequence learning. *Memory & Cognition*, **35**, 156-165.

FARRAND, P., & JONES, D. (1996). Direction of report in spatial and verbal serial short term memory. *Quarterly Journal of Experimental Psychology*, **49A**, 140-158.

HENDRY, L., & TEHAN, G. (2005). An item/order trade-off explanation of word length and generation effects. *Memory*, **13**, 364-371.

HULME, C., NEATH, I., STUART, G., SHOSTAK, L., SURPRENANT, A. M., & BROWN, G. D. A. (2006). The distinctiveness of the word-length effect. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **32**, 586-594.

HULME, C., ROODENRYS, S., SCHWEICKERT, R., BROWN, G. D. A., MARTIN, S., & STUART, G. (1997). Word-frequency effects on short-term memory tasks: Evidence for a redintegration process in immediate serial recall. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **23**, 1217-1232.

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.

JALBERT, A., NEATH, I., BIRETA, T. J., & SURPRENANT, A. M. (2009). *When does length cause the word length effect?* Manuscript submitted for publication.

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

LEWANDOWSKY, S., BROWN, G. D. A., & THOMAS, J. L. (2009). Traveling economically through memory space: Characterizing output order in memory for serial order. *Memory & Cognition*, **37**, 181-193.

LEWANDOWSKY, S., & FARRELL, S. (2008). Short-term memory: New data and a model. In B. H. Ross (Ed.), *The psychology of learning and motivation* (Vol. 49, pp. 1-48). San Diego: Elsevier.

LEWANDOWSKY, S., & MURDOCK, B. B., JR. (1989). Memory for serial order. *Psychological Review*, **96**, 25-57.

LEWANDOWSKY, S., NIMMO, S., & BROWN, G. D. A. (2008). When temporal isolation benefits memory for serial order. *Journal of Memory & Language*, **58**, 415-428.

LI, S.-C., & LEWANDOWSKY, S. (1993). Intralist distractors and recall direction: Constraints on models of memory for serial order. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **19**, 895-908.

LI, S.-C., & LEWANDOWSKY, S. (1995). Forward and backward recall: Different retrieval processes. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **21**, 837-847.

MURRAY, D. J. (1968). Articulation and acoustic confusability in short-term memory. *Journal of Experimental Psychology*, **78**, 679-684.

NAIRNE, J. S. (1990). A feature model of immediate memory. *Memory & Cognition*, **18**, 251-269.

NAIRNE, J. S., NEATH, I., & SERRA, M. (1997). Proactive interference plays a role in the word-length effect. *Psychonomic Bulletin & Review*, **4**, 541-545.

NEALE, K., & TEHAN, G. (2007). Age and redintegration in immediate memory and their relationship to task difficulty. *Memory & Cognition*, **35**, 1940-1953.

NEATH, I. (2000). Modeling the effects of irrelevant speech on memory. *Psychonomic Bulletin & Review*, **7**, 403-423.

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 applications of a local distinctiveness model of memory. In B. H. Ross (Ed.), *The psychology of learning and motivation* (pp. 201-243). San Diego: Academic Press.

NEATH, I., FARLEY, L. A., & SURPRENANT, A. M. (2003). Directly assessing the relationship between irrelevant speech and articulatory suppression. *Quarterly Journal of Experimental Psychology*, **56A**, 1269-1278.

NEATH, I., GUÉRARD, K., JALBERT, A., BIRETA, T. J., & SURPRENANT, A. M. (2009). Irrelevant speech effects and statistical learning. *Quarterly Journal of Experimental Psychology*, **62**, 1551-1559.

NEATH, I., & NAIRNE, J. S. (1995). Word-length effects in immediate memory: Overwriting trace decay theory. *Psychonomic Bulletin & Review*, **2**, 429-441.

NEATH, I., & SURPRENANT, A. M. (2003). *Human memory* (2nd ed.). Belmont, CA: Wadsworth.

PAGE, M. P. A., & NORRIS, D. [G.] (1998). The primacy model: A new model of immediate serial recall. *Psychological Review*, **105**, 761-781.

PAGE, M. P. A., & NORRIS, D. G. (2003). The irrelevant sound effect: What needs modelling, and a tentative model. *Quarterly Journal of Experimental Psychology*, **56A**, 1289-1300.

ROSEN, V. M., & ENGLE, R. W. (1997). Forward and backward serial recall. *Intelligence*, **25**, 37-47.

SURPRENANT, A. M., NEATH, I., & BROWN, G. D. A. (2006). Modeling age-related differences in immediate memory using SIMPLE. *Journal of Memory & Language*, **55**, 572-586.

SURPRENANT, A. M., NEATH, I., & LECOMPTE, D. C. (1999). Irrelevant speech, phonological similarity, and presentation modality. *Memory*, **7**, 405-420.

SZMALEC, A., VANDIERENDONCK, A., & KEMPS, E. (2005). Response selection involves executive control: Evidence from the selective interference paradigm. *Memory & Cognition*, **33**, 531-541.

TEHAN, G., HENDRY, L., & KOCINSKI, D. (2001). Word length and phonological similarity effects in simple, complex and delayed serial recall task: Implications for working memory. *Memory*, **9**, 333-348.

TEHAN, G., & MILLS, K. (2007). Working memory and short-term memory storage: What does backward recall tell us? In N. Osaka, R. Logie, & M. D'Esposito (Eds.), *The cognitive neuroscience of working memory* (pp. 153-163). Oxford: Oxford University Press.

TEHAN, G., & TOLAN, G. A. (2007). Word length effects in long-term memory. *Journal of Memory & Language*, **56**, 35-48.

THOMAS, J. G., MILNER, H. R., & HABERLANDT, K. F. (2003). Forward and backward recall: Different response time patterns, same retrieval order. *Psychological Science*, **14**, 169-174.

THORNDIKE, E. L., & LORGE, I. (1944). *The teacher's word book of 30,000 words*. New York: Columbia University, Teachers College Press.

WALKER, I., & HULME, C. (1999). Concrete words are easier to recall than abstract words: Evidence for a semantic contribution to short-term serial recall. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **25**, 1256-1271.

WATKINS, M. J. (1972). Locus of the modality effect in free recall. *Journal of Verbal Learning & Verbal Behavior*, **11**, 644-648.

WATKINS, M. J., WATKINS, O. C., & CROWDER, R. G. (1974). The modality effect in free and serial recall as a function of phonological similarity. *Journal of Verbal Learning & Verbal Behavior*, **13**, 430-447.

NOTES

1. A similar experiment with forward recall was conducted, but because the output method was different (spoken vs. written), making direct comparisons is problematic.

2. The developers of OSCAR state "we regard backward recall as being outside the scope of OSCAR" (Brown, Preece, & Hulme, 2000, p. 174). An older model not included in Lewandowsky and Farrell's (2008) list, TODAM (Lewandowsky & Murdock, 1989), does address backward recall. We postpone consideration of this model until the General Discussion.

3. The figures show data collapsed over serial position, both for clarity and because there were no theoretically significant interactions involving serial position. Supplemental data showing recall as a function of serial position for all experiments are available from I.N. or at http://memory.psych.mun.ca/pubs/reports.shtml.

4. The figures are collapsed over word length, both for clarity and because word length did not enter into any significant interactions. Supplemental data showing output times as a function of condition for all experiments are available from I.N. or at http://memory.psych.mun.ca/pubs/reports.shtml.

5. We acknowledge that much recent work points to a positional rather than a temporal dimension in tasks such as serial recall. However, the temporal version requires fewer estimated parameters, and both accounts make almost identical predictions in situations such as those under consideration here (see, e.g., Surprenant, Neath, & Brown, 2006).

6. A Web-based implementation of this version of SIMPLE is available at http://memory.psych.mun.ca/models/.

APPENDIX
Words Used As Stimuli in Experiment 1, Equated for Frequency, Concreteness, Imageability, and Familiarity

Word	Concreteness	Familiarity	Imageability	Frequency Kučera & Francis	Thorndike– Lorge (Written)
		Short (One-Syllable)			
AISLE	509	503	528	6	72
BEAM	502	476	539	21	127
DRAW	442	542	435	56	428
FLOOD	553	523	598	19	325
HOWL	434	447	536	4	72
JOKE	388	580	483	22	230
LICE	543	397	532	2	4
MINK	589	524	604	5	27
PAIN	426	569	502	88	541
PEAL	402	451	433	1	13
PINT	483	536	487	13	92
ROSE	608	556	623	86	801
SALE	364	555	422	44	403
THREAT	335	524	408	42	108
WRATH	304	466	377	9	51
Mean	458.8	509.93	500.47	27.87	219.6
		Long (Three-Syllable)			
ABUNDANT	351	524	443	9	50
ACCIDENT	419	564	518	33	399
APPROVAL	267	526	375	51	108
ARTICLE	479	533	421	68	550
AVENUE	539	529	564	46	320
FOREIGNER	492	499	516	4	92
HEXAGON	559	387	527	1	4
MUSICIAN	564	558	585	23	72
OCCASION	346	566	305	58	424
PARAGRAPH	493	559	482	12	72
RECITAL	476	468	495	8	27
SEDATIVE	459	423	459	1	13
SYMPATHY	278	501	402	36	228
TELEGRAPH	547	460	518	21	126
TELEPHONE	619	605	655	76	800
Mean	459.2	513.47	484.33	29.8	219

Note—All values obtained from the MRC Psycholinguistic Database (Coltheart, 1981; www.psy.uwa.edu.au/mrcdatabase/uwa_mrc.htm).