Memory & Cognition
1991, 19 (4), 387-400

The role of assembled phonology in reading comprehension

VERONIKA COLTHEART
Macquarie University, Sydney, New South Wales, Australia

and

S. E. AVONS, JACQUELINE MASTERSON, and VERONICA J. LAXON
City of London Polytechnic, London, England

The contribution of assembled phonology to phonological effects in reading comprehension was assessed. In Experiment 1, subjects judged the acceptability of sentences with regular, exception, and nonword homophone substitutions and orthographic controls. Significantly more errors occurred to sentences with regular-word homophones than to exception words, and error rates for nonword homophones were low and not significant. Experiment 2 showed that this was not due to differences in the sentence frames. In Experiment 3, the subjects judged as unacceptable those sentences containing an exception word that sounded correct when read according to spelling-to-sound rules. Significantly higher error rates occurred only for low-frequency exception words. Experiment 4 showed that task conditions affect semantic-categorization error rates for nonword homophones. These results indicate that both assembled and addressed phonology contribute to sentence and word comprehension, but the low error rate for nonwords suggests that an early lexical check may be applied.

It has been argued for at least 10 years, on the basis of both experimental and neuropsychological evidence, that the comprehension of familiar written words is not phonologically mediated (Ellis, Miller, Sin, 1983; Patterson, 1982). When we consider the role of phonological mediation in sentence reading and comprehension, the situation is less clear. Evidence from printed-sentence comprehension implicates phonological mediation, at least to some extent. Experiments on phrase evaluation (Baron, 1973) and subsequently on sentence evaluation (Doctor, 1978; Treiman, Freyd, & Baron, 1983) showed that skilled readers fail to reject orthographically incorrect sentences that sound acceptable, for example, *A beech has sand.*

Phonological mediation in reading comprehension could occur if the phonological form of a word is activated and then acts as the input to a representation in the cognitive-semantic system, which stores word meanings. A comprehensive model that represents various procedures by which printed words may be comprehended and read aloud was presented by Patterson (1986; see also Howard and Franklin, 1987, and Monsell, 1987). The model, illustrated in Figure 1, incorporates an orthographic-input lexicon that contains representations of all words the reader can recognize in print. These are connected to semantic representations in the cognitive-semantic system and to entries in a phonological-output lexicon.

According to this model, *addressed* phonological mediation occurs when a phonological output is activated directly from the orthographic-input lexicon and converted into a phonological-input code. This indirect procedure accounts for the erroneous comprehension of exception words by surface dyslexic patients who may define *bury* as "a fruit" and *bear* as "uncovered or undressed" (Patterson, Marshall, & Coltheart, 1985). Another form of phonological mediation depends not on stored lexical phonology but on the *assembly* of phonology based on the application of letter-sound correspondences and possibly other subword segments. Reliance on this procedure results in "regularization" errors in reading aloud, for example, *bear* → "beer," and definitions based on the incorrect phonology obtained through assembled phonology, that is, defining *bear* as "an alcoholic drink." These sorts of errors are also made by surface dyslexics (M. Coltheart, Masterson, Byng, Prior, & Riddoch, 1983). It may be noted that the surface dyslexics who rely excessively on assembled phonology in reading comprehension are at chance in discriminating regular homophones of comparable familiarity and are likely to confuse words and pseudohomophones, for example, *staik*.

We can now consider which form of phonological mediation causes the errors found in sentence evaluation. From evidence to date, it appears that phonological mediation certainly arises through addressed phonology. This must be so, since errors occur when the homophones are irregular in spelling-to-sound correspondence, and their

Support for this research was provided by an Inner London Education Authority Research Fellowship held by J. Masterson, and ARGS/ARC Grant A78715724 to V. Coltheart. S. E. Avons is now at Wollongong University, Wollongong, NSW, Australia. Correspondence should be addressed to V. Coltheart, School of Behavioural Sciences, Macquarie University, NSW 2109, Australia.

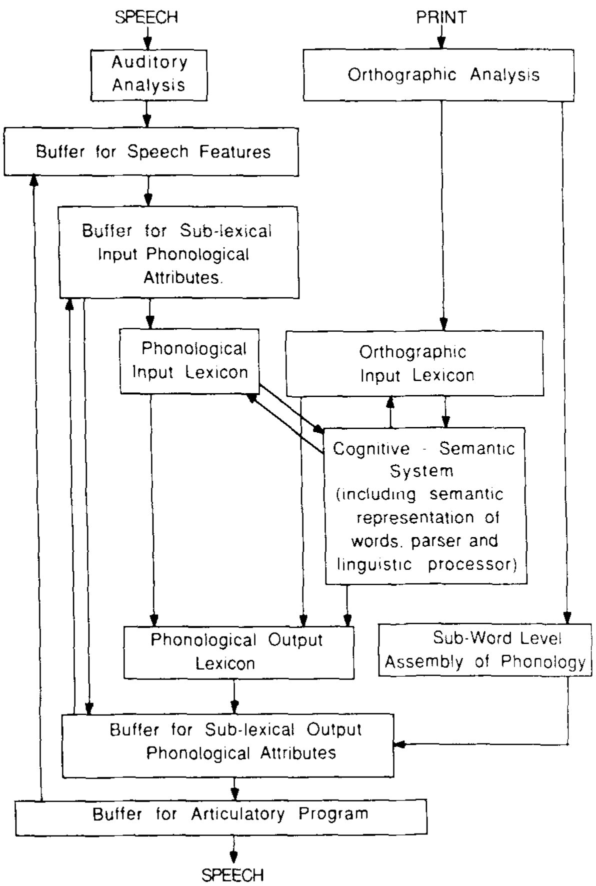

Figure 1. A model of reading and verbal short-term memory based on the models of Patterson (1986) and Monsell (1987).[1]

phonology must be addressed in the phonological-output lexicon (V. Coltheart, Laxon, Rickard, & Elton, 1988). Evidence concerning the contribution of assembled phonology is less clear. Treiman, Freyd, and Baron (1983) found that adults were more likely to accept as correct sentences containing an exception word that could be regularized to make the sentence sound correct, for example, *The are was quite polluted*. The control sentences used the correct regular word, *air* in the preceding example, in a sentence in which the irregular word (*are*) should have been used, for example, *Who air all these people*? Although the difference in error rates to these two types of sentences was significant, the effect did not generalize over stimuli. Thus, for this paradigm, assembled phonology has only a small observable effect on performance. This may be because the regularization produced by assembled phonology is overwhelmed by competition from addressed phonology or because of semantic or other constraints on acceptability arising orthographically.

In a variant of this paradigm, Waters, Seidenberg, and Bruck (1984) speculated whether the regularization that produces errors might in fact be confined to low-frequency exception words. They presented exception words and regular words as completions to a preceding incomplete sentence. Decisions for acceptable completions were slower and less accurate to exception words than to regular words, but this effect was confined to low-frequency words. Stimuli analyses were not reported, and so it is not known whether the regularity effect generalized over stimuli in this case either. Moreover, no attempts were made to ensure that the contexts provided for regular and exception words were comparable. Since a two-sentence context preceded the judgment to a single word, the Waters et al. task was rather unusual, and the experiments reported in this paper used more natural-sentence acceptability and semantic-categorization paradigms.

Another means of investigating assembled phonology is to use nonwords. When pseudohomophone substitutions were made in a sentence-acceptability paradigm (e.g., *They wate for the bus*) a significant false-positive error rate was observed (V. Coltheart et al., 1988). This indicates that assembled phonology is used in reading comprehension, since nonwords do not have lexical representation. However, the error rate was small, and sometimes this homophone effect was not found (V. Coltheart, Avons, & Trollope, 1990; Doctor, 1978). Thus, evidence for the involvement of assembled phonology in printed-sentence comprehension is relatively meager. In the experiments reported here, we sought to establish its role more conclusively, since an obligatory role for assembled phonology in printed-word comprehension has been claimed by Van Orden, Johnston, and Hale (1988).

EXPERIMENT 1

Earlier research has not indicated the extent to which addressed and assembled phonology contribute to phonological mediation in reading comprehension, since none of the experiments contrasted homophones that are regular in spelling-to-sound correspondence with those that are not.

Consequently, Experiment 1 presented a sentence-evaluation task with inappropriate homophones of three types: exception words, regular words, and nonword homophones. If false-positive errors to these unacceptable sentences arise only through addressed phonology, then errors to sentences with exception words should be as frequent as errors to those with regular homophones. If assembled phonology also contributes to phonological mediation, then errors to sentences with regular homophones should exceed those to sentences with exception-word homophones. Additionally, a homophone effect might be obtained with nonword homophones, as in at least one earlier study (V. Coltheart et al., 1988). As in earlier studies, orthographically matched control words and nonwords were also presented in sentences that *sounded* unacceptable.

Method

Subjects. Twenty-three undergraduate and postgraduate students and staff of the City of London Polytechnic served as subjects in Experiment 1 and were paid for their participation. One subject with a very high error rate was replaced.

Table 1
Characteristics of Target Words and Nonwords in Experiment 1

	Type of Target					
	Regular Homophone	Control Word	Irregular Homophone	Control Word	Nonword Homophone	Nonword Control
Mean frequency	92	90	97	96	94	89
Mean graphic similarity	593	614	584	588	590	583
Mean length (letters)	4.6	4.5	4.6	4.4	4.5	4.3

Stimulus materials. The target items in the experiment consisted of 20 regular homophones, 20 exception-word homophones (Venezky, 1970), and 20 homophonic nonwords. These word and nonword homophones are homophonic for speakers in Southern England. Two sets of 20 control words were selected for the homophones, and 20 control nonwords were devised for the homophonic nonwords. These 120 target words and nonwords were inserted into sentences that were semantically unacceptable. Sentences containing the homophones were phonologically acceptable; those containing the controls (words and nonwords) were not. Examples are shown below:

1. Regular homophone sentence: *She washed the window pain.*
2. Regular control sentence: *The pine of glass was broken.*
3. Exception-word homophone sentence: *The eagle sword in the sky.*
4. Exception control sentence: *The bird scared in the air.*
5. Nonword homophone sentence: *Her bloo dress was new.*
6. Nonword control sentence: *The sky is bloe today.*

Two sets of unacceptable sentences were devised: one set contained the homophones and controls assigned to sentence frames, as above; the other set reversed the assignments. Thus, the regular homophone sentence became *The pain of glass was broken*, and its control was *She washed the window pine*. Similarly the exception-word homophone, *sword*, now appeared in *The bird sword in the air*, and the control word, *scared*, was in *The eagle scared in the sky*. Nonword homophones and controls were similarly switched across their sentence frames. For sentences with exception-word homophones and controls, in six pairs of sentences, both homophones and controls were syntactically and semantically unacceptable; in five pairs, both homophones and controls were syntactically acceptable; in four pairs, the homophones were syntactically acceptable; and in five pairs, the control words were syntactically acceptable. For sentences with regular homophones and controls, the distribution of syntactically acceptable and unacceptable sentences was similar.

A separate set of 120 acceptable sentences was also devised. In these, the homophones and control words were used correctly. Acceptable sentences based on the nonword homophone and control targets used the correctly spelled word homophones of the nonwords (i.e., if the target nonword was *bair*, the acceptable sentence contained the word, *bear*, correctly used).

The target homophones and controls were matched for various characteristics such as word frequency (Hofland & Johansson, 1982) and length (number of letters). They were also matched in graphic similarity to the word that should have been used in the sentence (using Weber's, 1970, index of graphic similarity). Although values of Weber's index are reported, the homophones and controls are also matched if Van Orden's (1987) adjusted orthographic similarity (O.S.) scores are used. Finally, homophones and controls were matched, where possible, in the pattern of letters with ascenders and descenders. Mean graphic similarity values, word frequencies, and word lengths are presented in Table 1, and the target words and nonwords are listed in Appendix A. Word frequencies for the homophone mates of the regular and exception words were similar, except that one of the mates of an exception word (*would*) had a very high frequency.

Sentences varied in length from 4 to 8 words, but the mean length for each set of 20 was 5.6 words. Target words or nonwords occurred approximately equally often at the beginning, middle, or end of sentences.

Finally, a set of 16 comparable sentences that used different homophones and controls was devised as a practice set. It may be noted that most of the sentences were those used by V. Coltheart et al. (1988) and V. Coltheart et al. (1990).

Apparatus and Procedure. The procedure followed was the same as that used by V. Coltheart et al. (1988). A BBC microcomputer was used to display the sentences that were printed in lowercase in the middle of the screen. The subjects were instructed to judge whether each sentence that was presented was an appropriately spelled correct English sentence that made sense and were informed of the total number of sentences and the proportion correct. The subjects controlled presentation rate by means of the space bar. Two response keys (one each for "yes" and "no" responses) were located at opposite sides of the keyboard. Half of the subjects used the preferred hand for "yes" responses, and half used the nonpreferred hand. Sixteen practice sentences were presented before the main set. Presentation of the 240 sentences was randomized anew for each subject. Response times and errors were recorded automatically.

Results

Mean percentage errors and correct reaction times (RTs) on unacceptable and acceptable sentences are presented in Table 2.

Unacceptable sentences: Error data. The total errors made by each subject on each of the various types of unacceptable sentences were subjected to an analysis of variance (ANOVA) with type of target (exception, regular, and nonword) and homophony (homophone and control) as within-subjects factors. A second ANOVA, in which the random factor was stimuli and the fixed factors were type of target and homophony, was also performed.[2] Sub-

Table 2
Mean Percentage Errors and Mean Reaction Times (RTs) on Unacceptable and Acceptable Sentences in Experiment 1

	Homophone (Sound Right)		Control (Sound Wrong)	
	% Errors	RTs*	% Errors	RTs*
	Unacceptable Sentences			
Regular	27.8	1,982	10.5	2,052
Exception	15.9	1,841	4.78	1,946
Nonword	2.8	1,389	3.3	1,562
	Acceptable Sentences			
Regular	7.2	1,881	8.7	1,942
Exception	4.1	1,792	4.3	1,766
Correct homophone for nonword	4.1	1,771	3.5	1,723

*In milliseconds.

jects made more false-positive errors to sentences that sounded acceptable than to those that did not. This was indicated in the large effect of homophony [$F(1,22) = 40.7$, $MS_e = 2.97$, $p < .001$, for subjects, and $F(1,57) = 23.45$, $MS_e = 5.91$, $p < .001$, for stimuli].

There were also significant differences among the three types of target items [$F(2,44) = 50.4$, $MS_e = 2.37$, $p < .001$, for subjects, and $F(2,57) = 14.57$, $MS_e = 9.69$, $p < .001$, for stimuli]. The main effects must be interpreted, however, in the light of the highly significant interaction between homophony and type of target [$F(2,44) = 24.9$, $MS_e = 1.51$, $p < .001$, for subjects, and $F(2,57) = 7.08$, $MS_e = 5.91$, $p < .005$, for stimuli].

This interaction was investigated with a series of planned comparisons that used the Bonferroni procedure. These showed that the homophone effect was significant for both exception-word homophones [$t(66) = 5.32$, $p < .01$, for subjects, and $t(57) = 3.45$, $p < .01$, for stimuli] and for regular homophones [$t(66) = 8.35$, $p < .01$, for subjects, and $t(57) = 5.07$, $p < .01$, for stimuli]. There was no significant difference between homophones and controls for sentences with a nonword.

Errors to sentences with regular homophones exceeded those to sentences with exception-word homophones [$t(88) = 5.82$, $p < .01$, for subjects, and $t(114) = 3.06$, $p < .05$, for stimuli]. Error rate to both types of homophones exceeded error rate to nonword homophones. For the exception-word homophone-nonword contrast, the statistics were $t(88) = 6.35$, $p < .01$, for subjects, and $t(114) = 3.45$, $p < .05$, for stimuli. The regular homophone-nonword homophone comparison yielded $t(88) = 12.17$, $p < .001$, for subjects, and $t(114) = 6.51$, $p < .01$, for stimuli.

Unacceptable sentences: RT data. Again, ANOVAs were performed with type of target and homophony as fixed factors and subjects and stimuli as random factors. Sentences with inappropriate homophones were rejected more quickly than were sentences with control words [$F(1,22) = 11.36$, $MS_e = 41{,}090$, $p < .005$, for subjects, and $F(1,57) = 3.97$, $MS_e = 127{,}797$, $p < .05$, for stimuli].

There was a significant main effect of type of target [$F(2,44) = 54.4$, $MS_e = 68{,}108$, $p < .001$, for subjects, and $F(2,57) = 16.76$, $MS_e = 259{,}307$, $p < .001$, for stimuli]. Planned comparisons indicated that sentences with nonwords were rejected more rapidly than were sentences with real-word homophones. The comparison between sentences with regular word targets and nonword targets yielded $t(44) = 9.95$, $p < .01$, for subjects, and $t(57) = 5.68$, $p < .01$, for stimuli. Similarly, the comparison between exception word and nonword rejection times was significant in both subject and stimulus analyses [$t(44) = 7.68$, $p < .01$, and $t(57) = 3.82$, $p < .01$, respectively]. In either analysis, the interaction between homophony and type of target was not significant.

Acceptable sentences: Error data. ANOVAs comparable to those for unacceptable sentences were performed. Type of target yielded a significant main effect [$F(2,44) = 9.54$, $MS_e = .993$, $p < .001$, for subjects, and $F(2,57) = 3.22$, $MS_e = 3.569$, $p < .05$, for stimuli]. This effect was attributable to the higher error rate on acceptable sentences based on regular homophones and their controls. However, the differences in error rates between these and the other types were significant only in the subjects analysis [$t(44) = 3.56$, $p < .01$, for the comparison with exception-word targets, and $t(44) = 3.97$, $p < .01$, for the comparison with correct versions of nonwords]. No other effects were significant.

Acceptable sentences: RT data. There was no significant difference in RTs to sentences with homophones and to sentences with control words. Nor did homophony interact with the type of target. The main effect of type of target was, however, significant [$F(2,22) = 18.2$, $p < .001$, for subjects, and $F(2,57) = 5.39$, $p < .01$, for stimuli]. Decisions to sentence sets with regular homophones and controls were slower than to sentence sets with exception homophones and controls [$t(44) = 4.57$, $p < .001$, for subjects, and $t(57) = 2.57$, $p < .05$, for stimuli] and were also slower than decisions to sentences based on nonwords [$t(44) = 5.68$, $p < .01$, for subjects, and $t(57) = 3.05$, $p < .01$, for stimuli].

Taking the error and RT data together, it can be seen that homophones, when correctly used, posed no special difficulties to the subjects. They neither slowed decision times nor increased error rates. However, there seemed to be some differences in difficulty across the sentence sets, since both errors and RTs were increased for the sentence sets using regular homophones and their control words.

Discussion

The results of Experiment 1 replicate previous studies that have found a substantial phonological effect with sentences with inappropriate homophones. Skilled readers erroneously accepted these sentences as correct. They were also more likely to do so when the homophones were regular in spelling-to-sound correspondence than when they were exception words. This suggests that assembled phonology might underlie the phonological mediation that causes errors. Although assembled phonology might be partly involved, it cannot be wholly responsible for the errors that have been observed. Since exception-word homophones also yielded a substantial effect, as they have done in previous studies (V. Coltheart et al., 1988), addressed phonology must also contribute to the phonological encoding that causes errors.

Error rates on sentences containing nonwords were very low, both in this experiment and in earlier ones (V. Coltheart et al., 1988; Doctor, 1978), and a homophone effect was not obtained. Additionally, both in Experiment 1 and in the earlier studies cited above, correct rejections of sentences with a nonword are substantially faster (by about 400 msec) than are rejections of sentences containing only real words. This faster rejection suggests that nonwords may be detected at an early stage, before the sentence has been fully processed. There may be an orthographic check that leads to their rejection and that is mostly completed before assembled phonology can take place or before its products can be processed by the auditory word-recognition system.

As in previous studies, rejection times were not increased by an inappropriate homophone. Indeed, such sentences were significantly more quickly rejected than were the control sentences. This result has not been reported previously. However, there is a nonsignificant tendency for faster rejection of unacceptable homophone sentences in at least two earlier studies (V. Coltheart et al., 1988; Doctor, 1978). We can only speculate on the reasons for this result and suggest that the subject has clearer evidence about the error in the sentences with inappropriate homophones: It is obvious which is the inappropriate word and which word should have been present. Possibly the subject is less able to detect the anomaly in the control-word sentences and may spend more time checking alternative interpretations before rejecting the sentence.

Thus, Experiment 1 indicates that phonological mediation is the result of both addressed and assembled phonology. However, the fact that error rates to both regular-homophone sentences and their controls exceeded those to exception homophones and their controls might arise from differences in the *sentence frames* used for the two sets of sentences. This possibility was examined in Experiment 2.

EXPERIMENT 2

Given that the sentence frames for the exception and regular homophones were different, it is possible that they were different in plausibility, predictability, or in some other way that caused the increase in error rate to the regular-homophone set. If differences in the sentence frames used across the two sets of homophones caused the differences in error rates, then differential error rates should also be observed when the sentences are acceptable and contain appropriate homophones.

To investigate this possibility, we conducted an experiment in which the homophones in the previously unacceptable sentences were altered to make them acceptable. Only those sentences containing exception and regular homophones were altered. However, to maintain equal numbers of acceptable and unacceptable sentences, the previously acceptable sentences (containing regular and exception homophones) were also altered to make them unacceptable. Thus, the task conditions and subjects were similar to those of Experiment 1, the only change was in the acceptability of the critical target sentences.

Method

Subjects. Thirty-two undergraduates and graduate students of the City of London Polytechnic acted as subjects and were paid for their participation. They were allocated to List A or List B in a balanced order. One subject whose error rate was unusually high was replaced.

Stimulus materials. Two hundred forty sentences, half acceptable and half unacceptable, were presented. Previously unacceptable sentences containing exception and regular homophones were altered to make them acceptable.

Examples of unacceptable regular-homophone sentences from Experiment 1:
List A: She washed the window pane.
List B: The pane of glass shattered.

Examples of unacceptable exception-homophone sentences from Experiment 1:
List A: The eagle soared in the sky.
List B: The bird soared in the air.

Similarly, previously acceptable sentences were altered to form two new sets of unacceptable sentences.

Acceptable sentences from Experiment 1:
The pane in his leg hurt a lot.
The soared is very sharp.

The remaining sentences from Experiment 1, namely unacceptable control-word sentences and nonword sentences, along with acceptable control and nonword sentences, were left unaltered.

Apparatus and Procedure. These were identical to those used in Experiment 1.

Results

Mean percentage errors and correct RTs to acceptable and unacceptable sentences are shown in Table 3.

Acceptable sentences: Error data. An ANOVA was performed with list (A or B) as a between-subjects factor and type of homophone (in Experiment 1) as a within-subjects factor. This indicated no significant effect of type of homophone. The list factor was not significant and did not interact with type of homophone. Thus, error rates across the various sentence sets were not different when these sentences had appropriate homophones.

Acceptable sentences: RT data. An ANOVA indicated a significant effect of type of homophone [$F(2,60) = 8.05$, $MS_e = 12{,}798.5$, $p < .001$]. This effect was also significant in an ANOVA with stimulus items as the random factor [$F(2,114) = 3.86$, $MS_e = 45{,}656.2$, $p < .05$]. Planned comparisons with the use of the Bonferroni procedure indicated that the sentences based on regular and exception homophones were not significantly different. The significant effect was attributable to the faster responses to acceptable sentences based on nonwords. Neither the list factor nor its interaction with type of homophone was significant.

Table 3
Mean Percentage Errors and Mean Reaction Times (RTs) for Acceptable and Unacceptable Sentences in Experiment 2

Category of Homophone in Experiment 1	List A % Errors	List A RTs*	List B % Errors	List B RTs*
Acceptable Sentences				
Regular	10.7	1,757	7.8	1,674
Exception	7.2	1,689	8.2	1,616
Nonword	6.9	1,681	8.2	1,523
Unacceptable Sentences				
Regular	14.1	1,880	16.6	1,694
Exception	12.8	1,795	14.4	1,607
Nonword	6.9	1,516	4.7	1,405

*In milliseconds.

Unacceptable sentences: Error data. Error rates were not different between Lists A and B. Errors were different among types of homophone [$F(2,60) = 10.01$, $MS_e = 3.30$, $p < .001$]. This difference was attributable to the substantially lower error rate to sentences containing a nonword homophone [$t(60) = 4.41$, $p < .01$], whereas error rates to all-word sentences were not different between the two sets of homophones. These error rates of 13%-16.5% to all-word sentences and 4.6%-6.9% to nonword sentences were comparable to those found in Experiment 1 and to other previous experiments cited earlier. Performance on unacceptable control sentences was similar to that observed in Experiment 1.

Table 4 reports correlations between error rates on unacceptable sentences in Experiments 1 and 2 and word frequency and graphic similarity. Two sets of correlations were calculated for word frequency (after a logarithmic transformation). In the first set, the correlation was based on the word frequency of the inappropriate homophone in the sentence, for example, *pain*. The second set of correlations was based on the frequency of the word that should have been used, that is, *pane*. None of these correlations was significant.

Unacceptable sentences: RT data. An ANOVA indicated that the effect of type of homophone was highly significant [$F(2,60) = 54.64$, $MS_e = 16{,}754.6$, $p < .001$]. This effect was largely attributable to the fact that sentences with a nonword homophone were rejected significantly faster than those with real words [$t(60) = 10.11$, $p < .001$]. However, performances on the two sets of all-word sentences were also significantly different, with slower responding to the sentences containing unacceptable homophone mates of the regular homophones of Experiment 1 [$t(60) = 2.66$, $p < .05$]. Again, performances on Lists A and B were not significantly different, nor did list interact with type of homophone.

Discussion

In Experiment 2, we attempted to discover whether differences in difficulty or plausibility of the unacceptable sentence frames of Experiment 1 caused the differential error rates and the rejection RTs between sentences with regular- and exception-word homophones. When the sentences contained the acceptable homophone mates of those used in Experiment 1, there were no differences in either the accuracy or the latencies of acceptability judgments. Thus, there was no evidence to suggest that the sentence frames caused the differences found in Experiment 1. We therefore conclude that regularity of spelling-to-sound correspondence caused the increased error rates and RTs to sentences with regular homophones in Experiment 1.

We also note that in Experiment 2 the usual error rates and rejection RTs to unacceptable sentences were obtained. The fact that error rates across the two sets of all-word homophone sentences were not different is probably attributable to both sets having comparable numbers of regular- and exception-word homophones (almost all were regular homophones). Error rates to these unacceptable sentences did not differ greatly from those to exception words in Experiment 1. This is of note since the homophones that should have been used (e.g., *through* instead of *threw*) were higher in frequency than those that should have been used in Experiment 1 (e.g., *threw* instead of *through*).

The low nonsignificant error rates to nonword homophones were also replicated in Experiment 2. Such low error rates have previously been reported for nonword homophones (V. Coltheart et al., 1988; V. Coltheart et al., 1990; Doctor, 1978). An explanation on the basis of frequency for these low error rates can be discounted. The nonword homophones were foils for correct word homophones of higher frequency than the correct homophones for the inappropriate exception and regular homophones that were used in Experiment 1. This occurred because the correct homophones for nonwords were matched in frequency to the unacceptable regular- and exception-word homophones. However, in Experiment 2, the previously unacceptable homophones, for example, *through*, were now the correct words that should have been used in the unacceptable sentences. Thus, word frequency for correct homophones (which should have been used) was closely matched across the sentence sets in which inappropriate word or nonword homophones were used. Error rates to unacceptable sentences with word homophones were still substantially higher than error rates to sentences with nonword homophones. Thus, the low error rates to sentences with nonwords cannot be attributed to differences in frequency of the words that should have been used in the sentences.

Table 4
Correlations of Error Rate on Unacceptable Sentences in Experiments 1 and 2 with Word Frequency and Graphic Similarity

	Type of Homophone			
	Experiment 1		Experiment 2 Homophone Mate of:	
Correlation of Error Rate with:	Regular e.g., *pain*	Exception e.g., *sword*	Regular e.g., *pane*	Exception e.g., *soared*
Log frequency of presented homophone	.12	−.37	−.06	−.19
Log frequency of homophone that should have been used	−.15	.17	−.13	−.31
Graphic similarity of homophone to the appropriate word	.41	.23	−.40	−.06

Furthermore, as indicated in Table 4, there appeared to be no relationship between the word frequency of either the unacceptable or the correct homophone and the error rate in Experiments 1 and 2 for the word-frequency range used in these experiments. It is possible that a relationship between word frequency and error rate would be found to exist if a greater range of frequency were to be examined (as claimed by Van Orden, 1987).

The increase in error rate for unacceptable sentences with a regular homophone appears to arise from a contribution of assembled phonology. For these homophones, both procedures for deriving a phonological representation yield the same result. We can now consider a slightly different situation in which addressed phonology yields a word that sounds wrong, whereas assembled phonology yields a phonological representation that sounds correct in the same sentence. This was the situation investigated by Treiman et al. (1983). In such sentences, the detection of any significant false-positive rate must indicate an effect of assembled phonology. The absence of a significant false-positive error rate does not mean that assembled phonology has no contribution; it may merely indicate that its effects are masked by the competing addressed phonology.

EXPERIMENT 3

In Experiment 3, we sought to replicate and extend the findings of Treiman et al. (1983), who showed that the use of assembled phonology might cause people to accept sentences with inappropriate exception words that can be "regularized" to yield a phonologically acceptable sentence. In this experiment, the subjects were presented with a sentence-evaluation task in which unacceptable sentences contained either an exception word or a control word. The exception words were chosen so that, if assembled phonology were used, they would yield a phonological representation that sounded correct in the sentence. Control words that were matched in frequency and, as far as possible, in orthographic similarity were chosen to always provide a phonologically unacceptable word in the same sentences.

The suggestion of Waters et al. (1984) that assembled phonology might only be used for lower frequency exception words was also tested. Word frequency was manipulated so that half of the exception words and the controls were of high frequency and half were of low frequency. If Waters et al. (1984) are correct, then the phonological effects in this task should be increased in, and possibly confined to, sentences containing low-frequency exception words. Since the intention was to replicate and extend the Treiman et al. (1983) findings, we used as many of their target exception words and sentences as possible. However, the control words were selected so that they matched the exception words in frequency, since that was a manipulated variable. They were also selected so that they would be inappropriate in the same sentence frame used for the exception word to which the control word was matched.

Method

Subjects. Twenty-four undergraduate and postgraduate students and staff of the City of London Polytechnic served as subjects in Experiment 3 and were paid for their participation.

Stimulus materials. The target items in the experiment consisted of 16 high-frequency and 16 low-frequency exception words. Control words were devised for the high- and low-frequency words according to criteria similar to that for Experiment 1. However, for high-frequency exception words, it was not possible to select control words that matched as closely in graphic similarity as did the control words in Experiment 1. Mean graphic similarity values, word frequencies, and word lengths are presented in Table 5.

As in Experiment 1, two sets of sentence frames were devised for the target and control items. One set consisted of 64 anomalous sentences. Sixteen sentences contained the low-frequency exception word (e.g., *He used to height his teacher*) that sounded correct if the target word was regularized, and 16 sentences contained the low-frequency exception-word controls (e.g., *The opposite of love is heat*). The other 32 anomalous sentences contained the high-frequency exception words and their controls.

In the set of 64 acceptable sentences, the exception words and their controls were used correctly (e.g., *What is the height of the building*? and *He was sick of the heat*). It should be noted that the target words and sentences included as many items as possible from the Treiman et al. (1983) spelling-sound rule task. Words unfamiliar in British usage or pronounced differently were excluded and additional words were selected to make up sets of high- and low-frequency words that sounded like other words when pronounced according to regular letter-sound correspondences.

Again, two presentation lists were prepared. List A and List B both contained the 64 acceptable sentences referred to above, but the 64 anomalous sentences were different across the two lists in that the target and control items were switched within the sentence frames across Lists A and B. A list of target words is to be found in Appendix B. Sixteen comparable sentences were devised for use as a practice set.

As in Experiment 1, the eight sets of sentences (anomalous and correct low-frequency exception words and controls, and anomalous and correct high-frequency exception words and controls) were matched for length and position of the target items within the sentences.

Apparatus and Procedure. The apparatus and procedure used were the same as that for Experiment 1.

Table 5
Characteristics of Target Words Used in Experiment 3

	High Frequency		Low Frequency	
	Exception Word	Control Word	Exception Word	Control Word
Mean frequency	1,001	855	26.6	27.1
Mean graphic similarity	639.1	456.3	636.4	585.0
Mean length (letters)	4.1	4.4	4.4	4.6

Results

Mean percentage errors and correct RTs on acceptable and unacceptable sentences are presented in Table 6.

Unacceptable sentences: Error data. An ANOVA was performed with word frequency (high, low) and type of target word (exception, control) as within-subjects factors. An ANOVA with stimuli as the random factor was also carried out.

Neither word frequency nor type of target word yielded significant main effects. However, there was a significant interaction between word frequency and type of target word [$F(1,23) = 4.22$, $MS_e = .891$, $p < .05$]. Simple main-effects tests indicated that for low-frequency words, significantly more errors occurred for sentences with an exception word than for sentences with a control word [$F(1,46) = 6.58$, $p < .01$]. For high-frequency target words, error rates did not differ significantly for the two types of sentences. However, in the stimuli analyses, no effects were significant (with F values < 2.0 and $p > .18$).

Unacceptable sentences: RT data. An ANOVA comparable to that for the error data was performed. This yielded a significant effect of frequency, with slower responses to sentences with high-frequency target words than to sentences with low-frequency targets [$F(1,23) = 24.28$, $MS_e = 26{,}366$, $p < .001$]. Neither type of target word nor its interaction with frequency were significant. Since the variables of interest did not produce significant effects, the analyses over stimuli were not performed.

Acceptable sentences: Error data. ANOVAs comparable to those reported earlier indicated that significantly more errors occurred to sentences with low-frequency target words than to those with high-frequency targets [$F(1,23) = 41.01$, $MS_e = .9764$, $p < .001$, for subjects, and $F(1,30) = 8.50$, $MS_e = 7.06$, $p < .01$, for stimuli]. No other effects were significant in either analysis.

Acceptable sentences: RT data. An ANOVA indicated that correct decisions were faster for sentences with high-frequency targets than for sentences with low-frequency targets [$F(1,23) = 19.13$, $MS_e = 10{,}480$, $p < .001$]. No other effects were significant. Since the type of target word and its interaction with word frequency were not significant, analyses over stimuli were not performed.

Table 6
Mean Percentage Errors and Mean Reaction Times (RTs) for Unacceptable and Acceptable Sentences in Experiment 3

	Type of Target Word			
	Exception Word (Sounds Right If Regularized)		Control (Sounds Wrong)	
	% Errors	RTs*	% Errors	RTs*
	Unacceptable Sentences			
High frequency	9.9	2,145	9.4	2,263
Low frequency	12.2	2,035	6.8	2,045
	Acceptable Sentences			
High frequency	3.6	1,887	1.8	1,804
Low frequency	10.4	1,923	11.2	1,952

*In milliseconds.

Discussion

In Experiment 3, we attempted to investigate more closely whether assembled phonology can cause errors in printed-sentence comprehension. Given the previous speculations concerning the importance of word frequency, this variable was explicitly manipulated. There was some evidence that errors arose from the use of assembled phonology for low-frequency exception words only. A significant effect was obtained, but it did not generalize over stimuli. It may be noted that Treiman et al. (1983) also failed to obtain significant effects in stimuli analyses. Had there been a strong effect, it should also have been observed with acceptable sentences, as found by Waters et al. (1984) for low-frequency exception words. Regularization in these cases should lead to more rejection of acceptable sentences. Unlike Waters et al., we found no effects on acceptable sentences, not even for low-frequency exception words.

If phonology plays a part in the comprehension of these sentences, then addressed phonology must largely determine performance. When addressed and assembled phonology dictate different pronunciations, it appears that the reader is more likely to rely on the output from addressed phonology. Rarely, if at all, does assembled phonology produce an error.

Considering the results of Experiments 1 and 3 together, it can be concluded that phonological mediation does occur in printed-sentence comprehension. Comprehension is very clearly mediated by addressed phonology, as shown in Experiment 1 (and earlier research previously mentioned). Assembled phonology also makes a contribution to this effect. In Experiment 3, evidence for assembled phonology was observed only for low-frequency exception words and then was not reliable over stimuli analyses.

The absence of a homophonic effect with nonwords in Experiments 1 and 2 is at variance with findings resulting from the use of a word-recognition paradigm. Recently, Van Orden (1987) and Van Orden, Johnston, and Hale (1988) reported substantial effects of phonological mediation in a single word-comprehension task for both word and nonword homophones. Van Orden (1987) presented subjects with printed categories, such as A FLOWER, followed by a target word that had to be classified as an instance of the category. Substantial error rates were found when negative instances were homophones of a category member, such as ROWS. Errors to orthographically matched controls, such as ROBS, were rare. This indicates a high level of phonological mediation in a single word-comprehension task that does not have the semantic and syntactic processing demands of the sentence-comprehension task.

Van Orden et al. (1988) included nonword homophones in two categorization experiments, for example, ARTICLE OF CLOTHING: SUTE and SURT. The error rate to homophonic nonwords was the same as that to words. They concluded that reading comprehension usually involves assembled phonology and semantic access, followed by an optional spelling check. This spelling check requires retrieval of the correct orthography of the target item, followed by comparison with the presented stimulus. These results contrast with those of Experiments 1 and 2, in which homophonic errors were found to words but not to nonwords. This difference may arise because different strategies of reading are adopted in these two rather different tasks or because of differences in the experimental conditions. The two tasks might prompt different strategies. Sentence comprehension requires much more complex semantic and syntactic integration. Since sentence-acceptability judgments are much longer than judgments of class membership, any deadline for decisions concerning lexicality is correspondingly lengthened. If so, then nonwords could be excluded, independent of their phonology, by a late lexicality decision. The second possibility is that differences in nonword responses were caused by procedural differences between the two kinds of experiment; this was examined in Experiment 4.

EXPERIMENT 4

There were two major differences between Experiment 4 and the previous experiments—instructions to the subjects and trial composition. The main difference in instructions was that in the sentence-reading task, the subjects were instructed to judge whether each sentence presented was an appropriately spelled, correct English sentence. (For example, "Half the sentences will be wrong because they contain one word which is incorrectly spelled or a word which makes the sentence meaningless. Respond YES if you think the sentence makes sense and is spelled correctly. Respond NO if you think the sentence does not make sense because it has an incorrect word or has a spelling mistake.") In contrast, Van Orden et al.'s (1988) subjects were not told to reject misspelled words. Van Orden (1984) did instruct subjects to reject misspellings and found a much reduced error rate of 8% to homophones. Thus, instructions have a dramatic effect on false alarms to homophones.

The trial structure in sentence-reading tasks were also different. The two paradigms were different with respect to the overall proportion of critical trials containing an inappropriate homophone or spelling control. For sentence acceptability, half of the total trials contained critical items in which a related word or nonword substitution had been made, and all the negative trials were critical trials. In contrast, the Van Orden paradigm included only 40 critical trials in a total of 200. The other trials consisted of 60 that used negative filler items that were not orthographically or phonologically related to category exemplars and 100 positive-category members. These differences in trial structure could influence subject strategies leading to different false-positive rates for the pseudohomophones and the word substitutions.

We investigated the possibility that procedural differences between Van Orden's studies and the sentence-acceptability paradigm were responsible for the discrepancies in word and nonword homophonic errors. Two versions of the semantic categorization task were constructed, one a replication of Van Orden et al.'s (1988) Experiment 1, the other a variant with fewer trials, in which word substitutions had been made in all of the negative trials. In addition, in this second condition, the subjects were given instructions alerting them to the possibility of misspellings in the word targets.

Method

Subjects. Forty-two adults, all students or nonteaching staff of the City of London Polytechnic, served as subjects. They were allocated by turns to one of the two conditions.

Stimulus materials. Positive instances and critical target items were category exemplars and nonexemplars taken from 20 categories of the Uyeda & Mandler (1980) norms. There were 40 critical target items, all of them nonexemplars, made up of four kinds: (1) words that were homophones of exemplars, (2) spelling control words that were orthographically similar to exemplars but not to homophones, (3) homophonic nonwords, and (4) spelling control nonwords. These 40 critical target items and their respective categories were those used by Van Orden et al. (1988).[3]

The positive instances were chosen to fulfill several criteria used by Van Orden et al. (1988). From each category used for the target items, positive instances were chosen that had category typicality values both above and below the exemplar corresponding to the critical target items for that category. Sixty negative instances (in Condition 1) were words matched for length and frequency to exemplars from each category but unrelated to the categories and category members. The experiment was controlled by a BBC Master microcomputer linked to a Hantarex low-persistence monitor. Responses were made by the W and P keys of the keyboard, which were labeled NO and YES.

Procedure. The subjects were tested individually. They read written instructions that explained the task, and then they carried out the practice set of 40 trials. After further instructions, they then completed the main experimental task, which consisted of 80 trials for the short condition and 200 trials for the long condition.

Instructions varied across conditions. In the long condition, the subjects were told simply to respond YES if the word was an exemplar of the preceding category, otherwise NO. In contrast, the subjects in the short condition were told to press the YES key if the word was an instance of the category and was spelled correctly and to press the NO key if they thought the word was not an instance of a category or was wrongly spelled.

Each trial began with the display of the word READY, and the subject pressed the space bar of the keyboard to continue. After a 500-msec interval, the category name was then displayed for 2 sec. This was immediately followed by a plus sign (+) that was displayed for 500 msec as a fixation point close to the center of the following target word. This fixation sign was displayed for 500 msec, and, after a dark interval of 20 msec, the target word was displayed and remained visible until the subject responded. The screen remained dark for a further second, and the READY signal then reappeared at the start of the next trial. All subjects were urged to respond as quickly as posssible and were told to look directly at the fixation point when it appeared. The entire experiment was completed in about 20-30 min, depending on the condition.

Results

The mean percentage errors and correct RTs for the critical items are shown in Table 7.

Negative instances: Error data. This shows that the overall error rate was low, at most only approaching 12%.

Table 7
Mean Percentage Errors and Mean Reaction Times (RTs) for Categorization of Negative Instances in Experiment 4

	Type of Negative Instance							
	Homophone		Control		Nonword Homophone		Nonword Control	
Condition	% Errors	RTs*	% Errors	RTs*	% Errors	RTs*	% Errors	RTs*
Long (200 trials)	9.5	757	5.7	707	11.4	755	1.9	733
Short (80 trials)	10.5	891	2.9	855	5.7	846	1.9	873

*In milliseconds.

An ANOVA was performed with condition (long, short) as a between-subjects factor and lexicality (word, nonword) and homophony (homophone, control) as within-subjects factors. A similar analysis with stimuli as the random factor was also carried out. There was a significant effect of homophony, with more categorization errors occurring to homophones than to controls [$F(1,40) = 21.2$, $MS_e = 0.76, p < .001$, for subjects, and $F(1,18) = 6.05$, $MS_e = 5.58, p < .05$, for stimuli]. In the analysis over subjects, there was also a significant effect of lexicality with error rates to words higher than to nonwords [$F(1,40) = 4.18, MS_e = 0.36, p < .05$]. Error rates were not significantly different across the long and short conditions. However, the three-way interaction of condition × lexicality × homophony was significant in the subjects analysis and approached significance in the stimuli analysis [$F(1,40) = 5.94, MS_e = 0.4, p < .05$, and $F(1,18) = 3.56, MS_e = 1.41, p < .08$, respectively].

The three-way interaction in the subjects analysis was examined in a number of simple main-effects tests. In the long condition, error rates to homophones were not significantly different for words and for nonwords, confirming Van Orden et al.'s (1988) findings. In the short condition, error rates were significantly lower for nonword homophones than they were for real-word homophones [$F(1,80) = 6.22, p < .025$], as in our Experiments 1 and 2. Error rates on control items were not significantly different in the short condition. However, in the long condition, error rates on word controls were significantly higher than they were on nonword controls [$F(1,80) = 3.98, p < .05$].

An examination of the homophone effect within each task, for words and nonwords separately, yielded the following results. In the long task, homophone errors exceeded control errors only for nonwords [$F(1,80) = 16.44, p < .001$]. In the short task, the homophone effect was significant only for words [$F(1,80) = 10.52$, $p < .005$].

Since error rates were so low and their distributions were not normal, a series of randomization tests proposed by Edgington (1987) were also performed.[4] They calculated the exact probablities of obtaining the above F values with 10,000 different random permutations of the data. The randomization tests confirmed the probabilities reported above, with the exception of the lexicality main effect, for which p was equal to .09.

Negative instances: RT data. The RT data in Table 7 are based on adjusted means for each subject. Individual RTs for each subject were discarded if they were more than two standard deviations from the mean correct RT. A three-way ANOVA indicated significantly longer RTs in the short condition [$F(1,40) = 4.74, MS_e = 145{,}540.4$, $p < .05$]. The effect of homophony approached significance [$F(1,40) = 3.32, MS_e = 5{,}059.9, p < .1$]. There was a significant interaction between lexicality and homophony [$F(1,40) = 7.69, MS_e = 2{,}730.4, p < .01$]. This interaction reflects the fact that for homophones, words were more slowly rejected than were nonwords, whereas for control items, words were rejected faster than nonwords, but in neither case was the difference significant.

Discussion

Considering the results of Experiments 1 and 3 together, it can be concluded that phonological mediation does occur in printed-sentence comprehension. It is very clearly mediated by addressed phonology, as shown in Experiment 1 (and earlier research previously mentioned). Assembled phonology also makes a contribution to this effect. In Experiment 3, evidence for assembled phonology was observed only for low-frequency exception words and then was not reliable over stimuli analyses.

Error rates in Experiment 4 were considerably lower than in the experiments of Van Orden et al. (1988). However, overall error rates were not different when subjects were alerted to misspellings and when all negative instances were orthographically similar to positive instances. Consequently, these differences in procedure do not increase accuracy overall. Nonetheless, some differences in the pattern of errors were found between the short and long conditions. A significant nonword homophone effect was obtained in the long condition, replicating Van Orden et al.'s (1988) results. Van Orden et al.'s homophone effect with words was not replicated. This may be due to the high error rate for control words, which was largely caused by a high error rate to BEER, the control for BEAR.

In the short condition, a homophone effect was obtained *only* with words. Error rates on nonword homophones were lower and not significantly different from error rates on nonword controls. These results are more comparable to those obtained with nonwords in the sentence-evaluation task, that is, in Experiment 1 and other previous studies. Given that the size of the homophone effect can be manipulated by both the nature of the instructions and the proportion of orthographically and phonologically related negative instances, we can question Van Orden et al.'s claim of an obligatory role of assembled phonology in reading comprehension. Even the necessity of using addressed phonology in this task can be queried, since a significant homophone effect with words was obtained in only one of the conditions.

If we compare our results to those of Van Orden et al. (1988) in more detail, it is immediately apparent that error rates to homophones were much lower in our study, even in the conditions that duplicated theirs: 9.5% and 11.4% to words and nonwords, respectively, compared with their 21% for both types of items. One possible reason for this difference might be that our subjects were simply slower and more careful than theirs. This was not the case, however, since in our study, correct rejection times were quite a lot faster—708-757 msec, compared with RTs of 936-979 msec in their experiment.

Another possibility is that the subjects in the Van Orden et al. (1988) study were less well acquainted with the appropriate spellings of the homophones used in the categorization experiment. This is indicated by the fact that in an *untimed* lexical-decision task (Van Orden et al., 1988), a separate group of subjects from the same pool of high school students erroneously judged 15% of the nonwords to be words. If this reflects their inadequate knowledge of the target words and nonwords, then the extent of genuine phonological mediation is in fact much lower in their experiment, namely, approximately 6%-7%. If this is the true extent of phonological mediation, it is not strong evidence for the claim that printed-word comprehension is invariably phonologically mediated.

There is another disquieting feature about the categorization-error data of Experiment 4. Errors are very unevenly distributed over items. In fact, the majority of them attracted few errors—the errors predominantly occurred to words and nonwords in which EE was replaced by EA, or vice versa. Thus, there were high error rates to JEAP, SLEAT, DEAR, STEEL, and so forth. The only other homophone to attract a relatively high error rate was PLAIN (instead of PLANE). Furthermore, an orthographic control involving an EE/EA substitution also attracted a high error rate (BEAR/BEER). We do not know whether Van Orden et al. (1988) found the same uneven distribution of errors, but note that in Van Orden's (1991) proofreading experiments, errors were also very high for nonword homophones involving an EA/EE substitution, for example, GREAN, SNEEK, SPEER. This specific pattern of errors confined to rather few items is not characteristic of errors found in the sentence-reading task, in which errors are rather more evenly distributed across words.

If errors are largely found for just a few items, this raises questions about the generality of the relationship between error rate and word frequency reported by Van Orden (1987) and Van Orden et al. (1988). This relationship is between error rate and the frequency of the word that should have been used. For example, the error rate to STEAL as a category member of METAL was positively related to the word frequency of STEEL. Just a few items may be producing this effect.

We have failed to find significant correlations between the error rate and the word frequency of either the (inappropriate) presented word or the word that should have been used (V. Coltheart et al., 1988). Similar correlations were calculated for the homophones used inappropriately in Experiments 1 and 2, along with correlations of error rate with graphic similarity to the appropriate word (reported in Table 6), and, again, none of the correlations was significant.

CONCLUSIONS

Experiment 1 shows that false-positive errors to homophone substitutions in sentence comprehension are more likely to arise from regular than from irregular homophones. Possible confounding effects of homophone frequency and visual confusion of the homophones for the target words can be ruled out, since the appropriate controls were included. Another possible explanation is that the sentences used to embed the two kinds of homophones varied in difficulty. Experiment 2 checked this possibility by measuring acceptability judgments on correct versions of the sentences in which no homophone substitutions were made. No differences were found between the sentences used as frames for regular homophones and those used for irregular homophones. These results suggest that there are differences in the processing of regular and exception words during sentence comprehension. This effect appears to be analogous to the word-regularity effect observed in word naming and in the lexical-decision task (e.g., Waters et al., 1984).

Homophone confusion errors should arise if access to the meaning of a word is mediated by its phonology. The excess of confusion errors with regular words would be expected if phonological descriptions are generated more quickly for regular words, without a corresponding increase in the speed of any competing process such as direct orthographic access to meaning. Dual-process theory (M. Coltheart, 1978) makes this prediction because the advantage of regular words rises through the use of a nonlexical procedure, which we have here referred to as assembled phonology.

Our data are consistent with a model of sentence comprehension in which access to semantics is sometimes, but not invariably, phonologically mediated. In Experiment 3, following Treiman et al. (1984), we used sentences with exception-word substitutions that would be phonologically acceptable if they were given a regularized pronunciation. The incidence of false-positive errors from regularization was small and significant only in analyses over subjects. One possibility is that comprehension here was mediated entirely by orthography. This seems unlikely in view of the high incidence of homophonic errors we reported in our other experiments. Perhaps a more satisfactory explanation is that phonological mediation does occur, but only the correct exception-word phonology is generated, and incorrect regularizations are excluded. If this interpretation is correct, it suggests that assembled phonology may speed up phonological recoding of a regular word for which the addressed and assembled pronunciations are compatible, leading to an increase in phonologically mediated access. But assembled phonology rarely overrules competing addressed phonology in the case of exception words. Certainly for good readers who pronounce exception words correctly, addressed phonology must take precedence over assembled phonology. In this paradigm, it appears to do so even with a sentential context that favors the competing regularized pronunciation.

Experiments 1 and 2 also considered the phonologically mediated, semantic interpretation of pseudohomophones. In agreement with many previous experiments on sentence comprehension, we found that false-positive error rates to pseudohomophones were smaller than error rates to word homophones, and the pseudohomophone false-positive error rates failed to reach significance. There are a number of possible explanations for this finding. One is that assembled phonology by itself is ineffective in providing access to semantics. On theoretical grounds, this seems unlikely, since we must then suppose that there are two types of internally generated phonology. Moreover, if assembled phonology could not access semantics, the phonological lexical-decision task (e.g., does *phocks* sound like a word?) would be impossible. Finally, the experimental evidence of Van Orden et al. (1988) and our Experiment 4 shows that in some circumstances, homophonic-confusion error rates can be as high for nonwords as they can for words.

The other explanations invoke orthographically based decisions to exclude sentences containing pseudohomophones. In the simplest account, pseudohomophones are detected by a spelling check that operates after phonologically mediated semantic activation. If pseudohomophone foils are more likely to be rejected as misspellings of the substituted word than are homophonic word foils, then the difference in error rates between words and pseudohomophones can be accounted for. One way this could arise is if subjects' knowledge of orthography is poorly specified. For instance, if they know that BARE and BEAR refer to an unclothed state and to an animal, but are uncertain which meaning corresponds to which spelling, then word homophone confusions will occur. However, these subjects should still be able to reject BAIR as a letter string that does not correspond to either meaning, resulting in a lower pseudohomophone error rate. Orthographic confusions between word homophones were not typical of our subjects. In a previous experiment, we assessed the orthographic knowledge that enables subjects to discriminate between homophones. Performance was high in an untimed test in which students (drawn from the same population) decided which of two alternative word homophones corresponded to a particular meaning.

A third possibility is that subjects rely on an independent lexical decision to detect and reject nonwords, including pseudohomophones. Although the evidence is inconclusive, two findings support this explanation. First, sentences containing nonwords are rejected much faster than all-word sentences, irrespective of their phonological status. Second, by using Van Orden's (1987) semantic-categorization task, we showed that changing the task structure and instructions altered the false-positive rate for pseudohomophones but not for words—pseudohomophone detection, therefore, seems to be under strategic control with the use of a mechanism that does not affect the error rate for words. The decision that rejects nonwords could be based on orthographic familiarity, the absence of orthographically based semantic activation, or (for parallel distributed processing, PDP, models) the orthographic-error score. What is important is to distinguish this from a spelling check that judges whether the orthography of an item is consistent with its current semantic interpretation.

Van Orden (1987) advocated a version of verification theory, in which access to meaning is always mediated phonologically followed by an orthographic check. In support, he notes that the error rate is related to the frequency of the substituted word in his semantic-categorization task. We have not observed this relationship in any of our experiments. The correlations between substituted word frequency and false-positive error rates are given in Table 6. We therefore find little support for Van Orden's account of word recognition. In our view, the data are best accounted for by a model in which access to semantics can take place by either an orthographic or a phonological route. Which meaning is accepted depends on the relative speed with which phonology is generated, the subject's strategy in using an orthographic check or lexical decision, and possibly higher level constraints provided by the context.

REFERENCES

Baron, J. (1973). Phonemic stage not necessary for reading. *Quarterly Journal of Experimental Psychology*, **25**, 241-246.

Coltheart, M. (1978). Lexical access in simple reading tasks. In G. Underwood (Ed.), *Strategies in information processing* (pp. 151-216). London: Academic Press.

Coltheart, M., Masterson, J., Byng, S., Prior, M., & Riddoch, J. (1983). Surface dyslexia. *Quarterly Journal of Experimental Psychology*, **35A**, 469-495.

Coltheart, V., Avons, S. E., & Trollope, J. (1990). Articulatory suppression and phonological codes in reading for meaning. *Quarterly Journal of Experimental Psychology*, **42A**, 375-399.

Coltheart, V., Laxon, V. J., Rickard, M., & Elton, C. (1988). Phonological recoding in reading for meaning by adults and children. *Journal of Experimental Psychology: Learning, Memory & Cognition*, **14**, 387-397.

Doctor, E. A. (1978). Studies in reading comprehension in children and adults. Unpublished doctoral dissertation, University of London, London.

Edgington, E. S. (1987). *Randomization tests* (2nd ed.). New York: Marcel Dekker.

Ellis, A. E., Miller, D., & Sin, G. (1983). Wernicke's aphasia and normal language processing: A case study in cognitive neuropsychology. *Cognition*, **15**, 111-144.

Hofland, K., & Johansson, S. (1982). *Word frequencies in British and American English*. London: Longman.

Howard, D., & Franklin, S. (1987). Three ways for understanding written words, and their use in two contrasting cases of surface dyslexia (together with an odd routine for making 'orthographic' errors in oral word production). In A. Allport, D. G. MacKay, W. Prinz, & E. Scheerer (Eds.), *Language perception and production: Relationships between listening, speaking, reading and writing* (pp. 340-366). London: Academic Press.

MONSELL, S. (1987). On the relation between lexical input and output pathways for speech. In A. Allport, D. G. Mackay, W. Prinz, & E. Scheerer (Eds.), *Language perception and production: Relationships between listening, speaking, reading and writing* (pp. 273-311). London: Academic Press.

PATTERSON, K. E. (1982). The relation between reading and phonological coding: Further neuropsychological observations. In A. W. Ellis (Ed.), *Normality and pathology in cognitive functions* (pp. 77-112). London: Erlbaum.

PATTERSON, K. (1986). Lexical but nonsemantic spelling? *Cognitive Neuropsychology*, **3**, 341-367.

PATTERSON, K. E., MARSHALL, J. C., & COLTHEART, M. (Eds.) (1985). *Surface dyslexia: Cognitive neuropsychological studies of phonological reading*. London: Erlbaum.

TREIMAN, R., FREYD, J. J., & BARON, J. (1983). Phonological recoding and use of spelling-sound rules in reading of sentences. *Journal of Verbal Learning & Verbal Behavior*, **22**, 682-700.

UYEDA, K. M., & MANDLER, G. (1980). Prototypicality norms for 28 semantic categories. *Behavior Research Methods & Instrumentation*, **12**, 587-595.

VAN ORDEN, G. C. (1984). *A ROWS is a /ROZ/ is a ROSE: Visual, phonological, developmental, and strategic influences upon lexical access*. Unpublished doctoral dissertation, University of California at San Diego, San Diego.

VAN ORDEN, G. C. (1987). A ROWS is a ROSE: Spelling, sound, and reading. *Memory & Cognition*, **15**, 181-198.

VAN ORDEN, G. C. (1991). Phonologic mediation is fundamental to reading. In D. Besner & G. Humphreys (Eds.), *Basic processes in reading: Visual word recognition* (pp. 77-103). Hillsdale, NJ: Erlbaum.

VAN ORDEN, G. C., JOHNSTON, J. C., HALE, B. L. (1988). Word identification in reading proceeds from spellng to sound to meaning. *Journal of Experimental Psychology: Learning, Memory & Cognition*, **14**, 371-386.

VENEZKY, R. L. (1970). The structure of English orthography. The Hague: Mouton.

WATERS, G. S., SEIDENBERG, M. S., & BRUCK, M. (1984). Children's and adults' use of spelling-sound information in three reading tasks. *Memory & Cognition*, **12**, 293-305.

WEBER, R. M. (1970). A linguistic analysis of first-grade reading errors. *Reading Research Quarterly*, **5**, 427-451.

NOTES

1. Monsell (1987), reviewing evidence on speech processing and production, along with verbal short-term memory, argued for the existence of speech input and output buffers and maintained that these, along with the pathways between them, constitute verbal short-term memory.

2. All of the significant effects reported in the Results section were still significant when the more conservative Greenhouse-Geiser adjustments to degrees of freedom are applied.

3. Van Orden et al. (1988) included BOLE/BOLB as a nonword homophone and control pair. Since BOLE is a word, this pair was replaced with BOAL/BOOL.

4. We thank Alan Taylor, who implemented the randomization test programs and ran these analyses. We are also grateful for the assistance he has provided with computing generally.

APPENDIX A
Target Words and Nonwords Used in Experiment 1

Regular Homophone	Control Word	Irregular Homophone	Control Word	Nonword Homophone	Control Nonword
meet (meat)	melt	wood (would)	word	wair (wear)	woar
fur (fir)	far	through (threw)	thought	teer (tear)	toar
fair (fare)	fire	seize (seas)	sense	bloo (blue)	bloe
here (hear)	hair	thrown (throne)	thorns	braik (brake)	broak
heel (heal)	hell	route (root)	rout	roze (rose)	rore
pain (pane)	pine	mown (moan)	moon	daize (days)	dores
style (stile)	stale	bury (berry)	busy	cawse (cause)	caise
raise (rays)	rise	pear (pair)	peer	cawd (cord)	cand
lays (laze)	lags	court (caught)	count	bair (bear)	bere
clause (claws)	claims	sword (soared)	scared	grone (grown)	grene
broach (brooch)	branch	suite (sweet)	seat	crooze (cruise)	craize
freeze (frees)	fries	heard (herd)	hard	weke (weak)	waik
fort (fought)	four	ball (bawl)	bail	chuse (chews)	choes
read (reed)	rend	earn (urn)	worn	dere (deer)	dair
eight (ate)	sight	heart (hart)	hurt	rele (real)	rool
board (bored)	bound	son (sun)	sin	pleeze (please)	plaize
taut (taught)	taunt	tow (toe)	ton	noaze (nose)	nuese
weighs (ways)	wages	stalk (stork)	stack	bete (beat)	bote
flew (flue)	flow	none (nun)	nine	saur (saw)	sean
peak (peek)	peck	soul (sole)	soil	hawl (hall)	heet

Note—Words in parentheses are the homophones that would have been correct in the sentence.

APPENDIX B
Target Words Used In Experiment 3

High-Frequency Words		Low-Frequency Words	
Exception	Control	Exception	Control
are	her	bear	bean
one	our	none	noun
were	there	height	heat
come	came	bread	brand
work	went	tour	tore
great	right	blows	blues
word	want	worm	wore
head	hand	sweat	swell
water	matter	shoe	shone
dead	deal	gauge	gouge
move	main	sew	sun
ready	really	plait	plant
use	once	plaid	planned
door	done	pear	pour
talk	took	pall	pail
child	called	fete	feast

(Manuscript received July 20, 1990;
revision accepted for publication January 30, 1991.)

Notices and Announcements

The Second International Conference on Music Perception and Cognition
University of California, Los Angeles
February 22-25, 1992

The Second International Conference on Music Perception and Cognition (2nd ICMPC) will be held, under the auspices of the Society for Music Perception and Cognition (SMPC), at the University of California, Los Angeles (UCLA) February 22-25, 1992. The Conference follows the highly successful first conference held in Kyoto, Japan, in October 1989. The 2nd ICMPC will provide an open forum for work in all areas of music perception, cognition, and related disciplines. There will be symposia of interest to a general audience and sessions of particular interest to specialists. Concerts and performance of all types of music, including non-Western, will be staged.

The conference welcomes a full range of research and scholarship, including experimental, theoretical, musicological, ethnomusicological, acoustical, physiological, and computational, which are directed toward the musical mind. In addition, performances of compositions that are based on these domains, or lecture-recitals including recorded tapes, are encouraged.

Abstracts of roughly 250 words should be submitted by September 30, 1991. Contributed papers are welcome. All papers should describe work that has not been presented at a conference or published in a journal. Proceedings will be published in consultation with the SMPC.

For further information, please contact Roger A. Kendall, 2nd ICMPC Secretariat, Department of Ethnomusicology and Systematic Musicology, 1642 Schoenberg Hall, UCLA, Los Angeles, CA 90024 (telephone, 213-206-6080; FAX, 213-206-6958).