Journal of Experimental Psychology:
Learning, Memory, and Cognition
1988, Vol. 14, No. 3, 371–386

0278-7393/88/$00.75

Word Identification in Reading Proceeds From Spelling to Sound to Meaning

Guy C. Van Orden
University of Colorado School of Medicine

James C. Johnston
National Aeronautics and Space Administration
Ames Research Center, Moffet Field, California

Benita L. Hale
University of Pennsylvania

Van Orden (1987) reported that false positive errors in a categorization task are elevated for homophonic foils (e.g., HARE for A PART OF THE HUMAN BODY). Two new experiments replicate this finding and extend it to nonword homophone foils (e.g., SUTE for AN ARTICLE OF CLOTHING). False positive errors to nonword homophone foils substantially exceed false positive errors to nonhomophonic nonword spelling controls, showing that the phonological characteristics of the nonword foils are critical. Because nonwords are not represented in the lexicon, this new result implicates computed phonological codes as a source of the categorization errors. Additionally, in each of two experiments, matched word and nonword homophones produced virtually identical error rates. If stimulus nonword homophones are viewed as extremely unfamiliar words, compared with the relatively familiar stimulus word homophones, then our failure to observe an effect of stimulus familiarity strengthens the case that phonological coding plays a role in the identification of all printed words. The fact that the results are obtained in a categorization task that requires reading for meaning (rather than a lexical decision task) makes it difficult to avoid the conclusion that phonological mediation plays a role in normal reading of text for meaning.

It has been a common assumption for many years that reading involves some form of representation analogous to the sound of words. Both the nature of these representations and their role in reading are, however, controversial. The nature of these representations is not our concern in the present article, and we will sidestep it by using the term *phonological representations* to refer to the superset that includes auditory, articulatory, phonetic, phonemic or more abstract phonological representations. Our focus here is the functional role in reading of words' phonological representations, whatever their exact nature.

Two possible functional roles can be distinguished: (a) Phonological representations may be used by the process of word identification, possibly in conjunction with orthographic representations (Barron, 1979, 1980; Besner, Davies, & Daniels, 1981; Coltheart, 1978; Coltheart, Davelaar, Jonasson, & Besner, 1977; Davelaar, Coltheart, Besner, & Jonasson, 1978; Gough, 1972; Gough & Cosky, 1977; McClelland & Rumelhart, 1981; Meyer & Ruddy, 1973; Naish, 1980; Perfetti, 1985; Rubenstein, Lewis, & Rubenstein, 1971; Rumelhart & McClelland, 1982; Spoehr & Smith, 1973; Stanovich & Bauer, 1978; Van Orden, 1987). (b) Phonological representations may be used after word identification, perhaps providing buffer storage for text integration processes (Baddeley, 1979; Baddeley, Eldridge, & Lewis, 1981; Baddeley & Lewis, 1981; Besner et al., 1981; Levy, 1978; Saffran & Marin, 1975). The second role now seems to be relatively well established. In contrast, the first role is hotly disputed. Many current theorists have explicitly argued against any use of phonological representations in word identification or have designed models that left no role for them (e.g., Aaronson & Ferres, 1983; Baron, 1973; Becker, 1976, 1980; Bower, 1970; Kleiman, 1975; Kolers, 1970; Paap, Newsome, McDonald, & Schvaneveldt, 1982; Smith, 1971).

Until recently, the best evidence that phonological representations mediate word identification came from experiments using the lexical decision task, in which subjects judge whether a letter string is a word (the reference language is usually English). In this task, subjects take more time to reject correctly pseudohomophone foils (e.g., DYME) than control foils (Coltheart et al., 1977; Rubenstein et al., 1971). It is reasonable to suppose that DYME is more difficult to classify as a nonword because it activates a phonological representation that, in turn, activates the lexical node for the word DIME. There is, however, a problem with extrapolating from this evidence to word identification generally.

We dedicate this article to our dear friend Benita Hale, who has died. We miss her.

This research was supported by a postdoctoral fellowship to Guy Van Orden, funded by the Center for Acoustical and Behavioral Research at AT&T Bell Laboratories. Preparation of the article was supported by a National Institute of Mental Health institutional postdoctoral fellowship through the Developmental Psychobiology Research Group at the University of Colorado Health Sciences Center, also to Guy Van Orden.

In addition to Marilyn Smith, Max Coltheart, Marcus Taft, and two anonymous reviewers, we thank Bruce Pennington, Alice Healy, Lori Wecker, Greg Matthews, Jim Lesser, Mary Jo Altom, Dave Turock, Steve Tell, Eileen Dahman, Saul Sternberg, and Ted Wright for their help with this project and their comments on the article.

Correspondence concerning this article should be addressed to Guy C. Van Orden, Department of Psychology, Arizona State University, Tempe, Arizona 85287.

Coltheart et al. (1977; see also Coltheart, 1978; Henderson, 1982; and McCusker, Hillinger, & Bias, 1981) noted that the slowing in response time to DYME occurs on "no" trials, and that "no" responses are typically much slower than "yes" responses. Thus, even if the prolongation of "no" response times to the nonword stimulus DYME were due to phonological mediation, that process might be too slow to play any substantial role in the identification of actual words. Consequently, this result may not be relevant to normal reading.

Van Orden (1987) avoided this problem of interpretation using a different task—categorization—and a different dependent measure—percentage of false positive responses. In the categorization task, subjects are presented with a category name (e.g., A PART OF THE HUMAN BODY) followed by a target word (e.g., TOOTH, HARE or SACK), which must be classified as to whether it belongs to the category. The key data come from target foils homophonic to category exemplars (e.g., target foil HARE for A PART OF THE HUMAN BODY). In Van Orden's (1987) Experiment 1, homophonic word foils like HARE were erroneously categorized as category exemplars on 18.5% of trials, versus only 3% for nonhomophonic control words equated for orthographic similarity. This paradigm avoids the problem noted previously because homophony effects show up on "yes" responses rather than the much slower "no" responses. Thus, Van Orden's (1987) results clearly support the hypothesis of phonological mediation in word identification.

Although Van Orden's (1987) results provide good evidence for an influence of phonological representations upon word identification, they do not specify the locus of this phonological influence. Two possible loci of phonological influence are illustrated in Figure 1. These loci are represented in the figure as separate processing routes, either of which might lead to a false positive categorization of the target foil HARE as A PART OF THE HUMAN BODY.

Route 1 is the classic phonological coding route: The stimulus HARE activates its orthographic representation, a temporary representation of spelling features (e.g., perhaps graphemes, as in Coltheart, 1978; or letter identities and positions, as in McClelland & Rumelhart, 1981; or some other, as yet unexplicated, spelling features). From this transient orthographic representation, the phonological representation /HAR/ is computed, in turn activating the lexical node "HAIR." Route 2 is a path producing the same homophonic error, but the phonological representation responsible is not computed from spelling features; rather, it is retrieved via direct lexical access of the orthographic representation. In the example, HARE's orthographic representation activates (without phonological mediation) the lexical node "HARE" and, as a consequence, activates the stored phonological representation /HÂR/. It is then this stored phonological representation /HÂR/ that activates the lexical node "HAIR." Although Route 2 is indirect, each of its component pathways is quite plausible. Consequently, before we can draw the conclusion that Route 1 provides the explanation for Van Orden's (1987) finding, Route 2 needs to be ruled out.

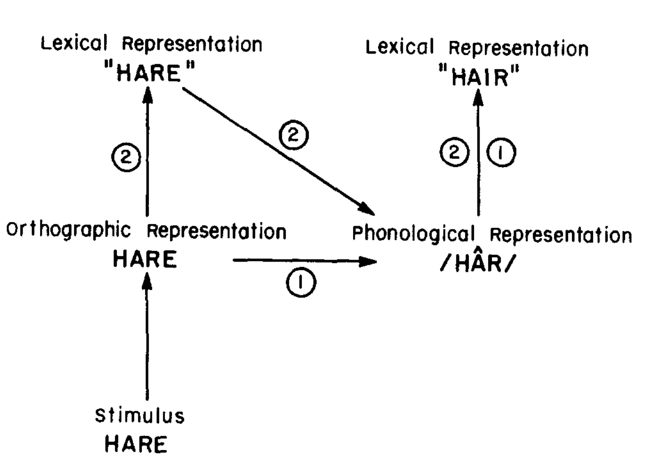

Figure 1. Phonological sources of activation at lexical entry "HAIR" when HARE is the categorization target. (Route 1 illustrates activation that arises from a computed phonological representation. Route 2 illustrates activation that comes from a retrieved phonological representation.)

Notice that Routes 1 and 2 yield phonological representations of very different natures. Routes 1's computed phonological representations are usually thought to be transient, in the same way that the orthographic representation from which they are computed is transient. But Route 2's retrieved phonological representations are permanent, in that they exist in the lexicon when inactive. Thus, although Figure 1 illustrates these two types of phonological representations conjunctively (to illustrate and emphasize their functional equivalence as sources of false positive errors to foils like HARE), they are not otherwise theoretically equivalent.

The present article attempts to strengthen the argument for computed phonological codes (Route 1), by extending the homophone effect to stimuli for which Route 2 is not possible, namely, *nonword homophone foils.* Consider the nonword target foil SUTE for the category AN ARTICLE OF CLOTHING. Route 2 is not viable because there is no word corresponding to S-U-T-E- to mediate the retrieval of the phonological representation /SO͞OT/ that could, in turn, activate the lexical representation of SUIT. It follows that if the lexical entry for SUIT is activated by a phonological representation, then that phonological representation must have been computed from the letters in SUTE (Route 1). Thus, conclusive evidence for the use of Route 1, the classic phonological coding route, could be found using SUTE as a target foil in a categorization task like that used by Van Orden (1987).

Route 2 (in Figure 1) could be a source of some false positive errors to nonword homophones like SUTE merely because most letter strings that are identical in sound are also similar in spelling. Accordingly, to determine whether miscategorizations of nonword homophones like SUTE can be attributed to their phonology, we use spelling control foils that are as similar in spelling to a corresponding category exemplar as each yoked homophone foil. For example, suppose that the nonword SURT and the nonword homophone SUTE are equally similar orthographically to SUIT. If miscategorizations are due only to orthographic similarity, then false positive errors to SURT and SUTE should occur with equal frequency. If, however, computed phonological codes (Route 1) are the main source of errors to SUTE, then fewer errors will be made to SURT.

Our design also incorporates a comparison between the proportion of categorization errors to nonword homophone foils (e.g., SUTE) and to word homophone foils (e.g., HARE) similar to those used by Van Orden (1987). This comparison should provide evidence about whether Route 2 in Figure 1 plays any role at all. We also include spelling controls for the word homophone foils. If more false positive errors are observed to word homophones like HARE than to yoked spelling controls like HARP, then we will have replicated the homophony effect observed by Van Orden (1987).

It would be tempting to compare error rates between word and nonword spelling controls, possibly concerning some hypothesis about effects of lexicality. This comparison, however, is contaminated by the lack of control over phonological similarity between these two stimulus types. These two sets of stimuli were matched closely for orthographic similarity to their respective, corresponding category exemplars. But it was impossible to match them also in relative phonological similarity to those same respective, corresponding category exemplars because this added constraint too severely restricted the pool of potential stimuli.

Experiment 1

Experiment 1 includes two variables: homophony (homophones vs. spelling controls) and lexicality (nonwords vs. words). Although all cells of this 2 × 2 design are included in the experiment, we are interested primarily in two specific comparisons. Both of these primary comparisons concern the loci of homophony effects in word identification. A comparison between the false positive error rate to nonword homophone foils and the false positive error rate to nonword spelling control foils tests for the influence of computed phonological codes (Route 1 in Figure 1) while controlling for the fact that homophonic letter-strings are usually orthographically similar. Additionally, a comparison between the false positive error rate to word homophone foils and the false positive error rate to nonword homophone foils tests for the effect of retrieved phonological representations (Route 2 in Figure 1). Experiment 1 also includes a comparison between correct "no" reaction times (RTs) to homophone and spelling control foils, as well as a comparison between false positive "yes" RTs to homophone foils and correct "yes" RTs to actual category exemplars.

Method

Subjects. The subjects, 30 students from New Jersey high schools, were paid for their participation. All had normal or corrected-to-normal vision.

Procedure. Subjects were seated in a sound-proof booth before a Hewlett Packard storage scope on which the stimuli were presented. The scope was controlled by a PDP11/23 computer. Each trial began with the presentation of the word READY. Subjects signaled that they were ready to begin by pressing a foot pedal. When the foot pedal was pressed, a category name appeared (e.g., AN ARTICLE OF CLOTHING). The category name was followed immediately by a plus sign (this fixation stimulus appeared in the exact center of the forthcoming target letter string). Subjects' instructions were to read the category name and then look directly at the plus sign. The plus sign was followed by a target letter string (e.g., SUTE). Upon presentation of the target, subjects responded "yes" if the target was an exemplar of the preceding category and "no" otherwise. Subjects responded by pulling a "yes" lever with their right hand or a "no" lever with their left hand.

Each session began with 40 practice trials; all of the practice targets were words and none were homophone foils. Subjects were instructed to use these trials to practice responding as quickly as possible while being accurate. The practice trials were followed by 200 experimental trials. Half of the trials included targets that were exemplars of their preceding categories. (This was true for both the practice and experimental trials.)

The practice trials were presented in the same order to all subjects. However, each subject was presented with a different random ordering of the experimental trials. The only condition on this ordering was that equal numbers of word and nonword homophones, and their respective spelling controls, appeared in the first and second halves of the experimental session.

An entire experimental session lasted about 45 min.

Viewing conditions. Stimuli were presented as light letters on a darker screen. No other light source was present within the booth, but minimal visibility was provided by ambient light sources. The timing of the stimulus presentation was as follows: The category name remained visible for 2 s, the plus sign for 500 ms, and the target until the subject responded.

Stimuli. A total of 240 targets were seen by each subject. Category names that appeared in the 40 practice trials did not appear in experimental trials. Each experimental category name (i.e., any category name that is followed in some trial by a homophonic foil) appeared in 6 "yes" trials and 6 "no" trials. Each target appeared once. The targets of interest were 10 nonword homophones (SUTE), 10 word homophones (HARE), and 20 respective nonword and word, yoked, spelling controls (SURT and HARP; a complete list of these targets appears in Appendix A with their category names).

Whenever possible, we constructed nonword homophone foils by transforming the spelling of the corresponding category exemplar in a way that mimics the orthographic relation that exists between the yoked word homophone and its corresponding category exemplar. For example, the nonword homophone SUTE is constructed from SUIT by deleting the third letter and adding an E to the end. This mimics the spelling transformation that would produce HARE (SUTE's yoked word homophone) from its corresponding category exemplar HAIR.

People's pronunciations of identical nonword letter strings do not always agree. To ensure that our nonword homophones were truly homophonic to their respective category exemplars, we tested a pool of candidate nonword homophones in a naming task. In this task, 10 judges (employees working at a variety of jobs at AT&T Bell Laboratories) read aloud rapidly a list of pronounceable nonwords. Candidate nonword homophones were randomly distributed throughout this list. Each judge saw a different random ordering of the list. The criteria used to select the eventual nonword homophone stimuli were that at least 9 out of 10 judges produced the homophonic pronunciation *and* any "mispronunciations" must be readily attributable to a misperception of the letters in the nonword homophone (e.g., pronouncing SUTE as SITE).

To verify the lexicality or nonlexicality of our homophone foils, we presented 20 high school student judges with a list of words and nonwords that included word and nonword homophone candidates. The judges were instructed to proceed through the list at their own pace indicating whether each letter string is a word or nonword. Each of the word homophone foils used in Experiment 1 was identified as a word by every judge; each of the nonword homophone foils used in Experiment 1 was identified as a nonword by at least 14 of the 20 judges.

Only one of the nonword homophones, BOLE, was actually identified as a word by six of our judges, and, in fact, it is an extremely rare word. We were, however, forced to use this stimulus because we could not find another nonword homophone that fit all other criteria. But, even including the "lexicality score" for BOLE, the nonword homophones in Experiment 1 generated a mean of only approximately 15% false positive lexicality judgments. (The nonword homophones in Experiment 2 also generated approximately 15% false positive lexicality judgments.) This is grossly different from the hit rate to the word homophone foils (100% for word homophones in Experiment 1 and approximately 96% for word homophones in Experiment 2).

The more similar HARE and HAIR are in spelling the greater the likelihood that HARE will be miscategorized as A PART OF THE HUMAN BODY (Van Orden, 1987). Therefore, each yoked pair of nonword and word homophone foils was matched closely in orthographic similarity to their corresponding category exemplars (e.g., SUTE is spelled as close to SUIT as HARE is to HAIR). The yoked spelling control of each nonword or word homophone foil was also matched for orthographic similarity to the same, respective, category exemplar. Mean orthographic similarity (OS, an estimate defined in Appendix B) for nonword homophone foils was .62 (*SD* = .128); for nonword spelling controls, .68 (*SD* = .109); for word homophone foils, .61 (*SD* = .080); and for word spelling controls, .64 (*SD* = .096).

We assume that matched OS scores between homophone foils and spelling control foils are adequate to control for orthographic similarity defined in terms of shared letter identities and positions, and shared letter pairs. Matching by OS scores does not, however, control for another *rule-governed* form of orthographic similarity. Taft (1982) has shown that nonwords that are *orthotactically* similar to real words (e.g., DREED is orthotactically similar to DREAD because EE and EA can be pronounced identically in some contexts) cause prolonged "no" response times in a lexical decision task. Unfortunately, given the other controls that we have included, it would have been impossible to add a further control for orthotactic similarity and still generate sufficient stimulus items. But, as we argue next, our experiment may not be compromised by this exclusion.

Van Orden (1984) has proposed an interactive-activation account of orthotactic similarity effects in which component spelling features activate phonological features, which in turn feed back activation to spelling features. By his logic, the spelling features corresponding to EE in DREED will activate the phonological features corresponding to the long /E/ sound, which will in turn feed back activation to the spelling features corresponding to both EE and EA. This top-down, phonologically mediated activation of the spelling features corresponding to EA can then add to the activation level of the lexical entry corresponding to DREAD. It is this activation of the lexical node "DREAD," caused by *phonologically mediated* activation of spelling features corresponding to EA, that causes a delayed "no" response time. Thus, it is reasonable to assume that phonological coding, like the phonological coding which the present experiments examine, is the source of orthotactic similarity effects.

The likelihood of a false positive categorization error to homophone foils (e.g., HARE) has been shown to increase as the word frequency (Kučera & Francis, 1967) of the corresponding category exemplar (e.g., HAIR) decreases (Van Orden, 1987). To avoid contamination of the comparison between nonword and word homophone foils, the category exemplars (e.g., SUIT and HAIR) corresponding to yoked pairs of nonword and word homophones (e.g., SUTE and HARE) were matched closely on the Kučera and Francis (1967) frequency counts. The mean log word frequency for category exemplars corresponding to nonword homophone foils was 1.19 (*SD* = .67) and for category exemplars corresponding to word homophone foils it was 1.32 (*SD* = .70).

All nonword and word homophone target foils were chosen to sound like category exemplars from the typicality norms of Uyeda and Mandler (1980). By using these norms, the typicality of category exemplars corresponding to pairs of nonword and word homophone foils was matched closely. For example, the word/nonword homophone pair HARE/SUTE was chosen such that the typicality of HAIR as A PART OF THE HUMAN BODY was close to the typicality of SUIT as AN ARTICLE OF CLOTHING. The mean typicality rating for the nonword homophones was 2.57 (*SD* = .888) and for the word homophones, 3.07 (*SD* = .846).

Six true category exemplar targets were chosen from among the sister exemplars of each true category exemplar (SUIT) corresponding to a homophone foil (SUTE). These category exemplar targets were chosen to be used in a comparison between false positive "yes" response times to foils like SUTE and correct "yes" response times to category exemplar targets like DRESS, SHOES, PANTS, BLOUSE, SOCKS, and VEST. The six category exemplar targets spanned a range of typicality extending both above and below the typicality rating of the category exemplar corresponding to a homophonic foil.

All remaining filler trials used category names from the norms of Uyeda and Mandler (1980) that did not appear in the trials of interest. The target category exemplars in filler "yes" trials also came from Uyeda and Mandler's norms. Filler "no" trials contained nonexemplar word targets that were *not* chosen to be systematically similar in orthography or phonology to actual category exemplars.

Results and Discussion

A single trial of interest, with response time under 150 ms (presumably an anticipation), was omitted from all analyses. The dependent measure for the error analyses was the percentage of key trials (trials that included homophone foils or spelling control foils) that resulted in a false positive error. (All of the error analyses in both of the experiments reported here were also computed using the arcsin transformation of the proportion of false positive errors [Winer, 1971], and virtually identical results were found.)

Separate tests were computed with both subjects and items as the random variable. For item analyses the sampling unit was either the stimulus quartet (for the "no"-latency analysis) composed of each pair of matched nonword and word homophone foils accompanied by their corresponding spelling controls (e.g., SUTE and HARE, SURT and HARP), or the stimulus pair appropriate to the planned comparison (for the error analysis). Of principal interest were two planned comparisons of false positive error rates between nonword homophones and their spelling controls, and between nonword homophones and word homophones. The balancing of stimulus properties of items was designed especially to ensure the soundness of these comparisons.

False positive error rates to key foils. Table 1 shows the percentage of false positive errors made to the key types of foils. The mean false positive error rate to nonword homophones like SUTE (21.3%) was significantly greater than the

Table 1

Percentage of False Positive Categorization Errors in Experiment 1 to Nonword and Word Homophone Foils Like SUTE *and* HARE*, Respectively, and Corresponding Yoked Spelling Controls like* SURT *and* HARP

Foil	*M*	*SE*	Control	*M*	*SE*
SUTE	21.3	2.7	SURT	3.0	1.3
HARE	21.8	2.6	HARP	2.3	1.0

false positive error rate to their yoked nonword spelling controls (3.0%), $t(9) = 3.00$, $p < .05$, for items and $t(29) = 7.49$, $p < .05$, for subjects. Thus, nonword homophones (SUTE) are much more likely to be mistaken for their sound-alike category exemplar (SUIT) than are their corresponding spelling controls (SURT). This result is strong evidence that computed phonological codes do influence word identification. We postpone further discussion of this finding until after Experiment 2. (Note peripherally that the error rate to nonword spelling controls is no greater than that to filler nonexemplar foils, 3.0%, words that were *not* chosen to be similar in either spelling or sound to category exemplars.)

The comparison of false positive error rates to nonword and word homophone foils showed virtually identical performance (21.3% and 21.8%, respectively), $t(9) = -.07$, for items, and $t(29) = -.18$, for subjects. We had not expected this result and will postpone full discussion of it until after Experiment 2, in which we attempt to replicate it. For now, it is worth noting the most obvious interpretation, that the source of miscategorizations of nonword homophone foils (presumably computed phonological codes) is also the source of miscategorizations of word homophone foils.

The comparison between the error rates to word homophone foils (21.8%) and their yoked spelling controls foils (2.3%) replicates the finding, reported in Van Orden (1987), of higher error rates to word homophone foils relative to their spelling controls. This previously established finding is not, however, crucial to the current hypotheses and will not be discussed here.

"No" latencies. The single trial that resulted in a response time of less than 150 ms was excluded from the "no" latency analysis. In addition, trials that resulted in response times greater than 2 s (a cutoff point well above 3 *SD*s from the grand mean) were excluded from this analysis (as well as the "yes" RT analysis). For any stimulus quartet's data (i.e., the data from yoked quartets like HARE/HARP/SUTE/SURT) to be included in the analysis of "no" response times, a subject must have correctly responded "no" to all four foils of that quartet. If a trial containing one of the foils in a yoked stimulus quartet resulted in a false positive error, then the three trials that contained the corresponding yoked foils were all excluded from the analysis of "no" RTs. (This yoking of trial outcomes ensured that each subject made a balanced contribution to all cells of the 2 × 2 factorial design consistent with the item sampling unit of carefully matched stimulus quartets.) The remaining data, from 56% of the total key trials and, approximately, 72% of the total key trials that resulted in correct "no" responses, were subjected to an analysis of variance (ANOVA). The reaction times for these correct "no" responses are shown in Table 2.

The apparent interaction between homophony and lexicality (see Table 2) is not statistically significant, $F(1, 29) = 2.41$, $p > .13$, for subjects, and $F(1, 9) = 1.83$, $p > .20$, for items.[1] Also, the mean "no" response time to nonword foils like SUTE and SURT (934 ms) was not significantly different from the mean "no" response time to word foils like HARE and HARP (935 ms; $F < 1$, for both subjects and items). However, correct mean "no" response times to homophonic foils like SUTE and HARE (958 ms) were significantly longer than correct mean "no" response times to spelling controls (912 ms), $F(1, 29) = 4.50$, $p < .05$, for subjects, and $F(1, 9) = 7.86$, $p < .05$, for items.

Table 2
Correct "No" Response Times (in Milliseconds) in Experiment 1 to Nonword and Word Homophone Foils Like SUTE and HARE, Respectively, and Corresponding Yoked Spelling Controls Like SURT and HARP

Foil	*M*	*SE*	Control	*M*	*SE*	*M*
SUTE	936	40	SURT	932	48	934 ms
HARE	979	52	HARP	892	39	935 ms
M	958		*M*	912		

The finding of longer "no" latencies to homophone foils than to spelling control foils would seem to replicate the results of the Meyer and Ruddy (1973) and Meyer and Gutschera (1975) categorization experiments, which were similar to Experiment 1. Inspection of the RT distributions that underlie the mean data showed, however, that the seemingly, generally delayed correct "no" responses to homophone

[1] The source of this interaction seems to be a subset of our nonword spelling controls that are pseudohomophones of real words, although they are not pseudohomophones of words that are exemplars of any of the categories that appeared in Experiment 1 (e.g., spelling control PARRIT is not homophonic to any exemplar of the category VEGETABLE, and the category name BIRD did not appear on any trial). The mean correct "no" response time for these pseudohomophonic spelling controls is 905 ms, but the mean correct "no" response time for nonword spelling controls that are not pseudohomophones is only 859 ms. Unfortunately, it was impossible to construct enough of these foils (within the other constraints of our experimental design) to allow a powerful test for the effect. If we had, however, found this effect to be significant, it would mean that the delay in response times that is caused by homophony cannot be avoided, even when a foil does not sound like a category exemplar. This would be consistent with the hypothesis that the effect of ambiguous phonology is localized in the process of word identification (see our, and Van Orden's, 1987, discussion of how spelling verification of stimuli with ambiguous phonology sometimes results in prolonged "no" RTs).

It is important to note, however, that the pseudohomophony of these spelling controls did not cause elevated false positive error rates. The error rate to these pseudohomophone spelling controls was 3.3%, almost identical to the error rate of 2.5% to other nonword spelling controls, and the error rate of 3.0% to filler foils that were not chosen to be systematically similar to exemplars in either spelling or sound. Thus, although they may prolong some process that is involved in categorization (presumably verification, a subprocess of word identification) they do not affect the outcome of the categorization decision. Consequently, it is not homophony per se that causes false positive categorization errors, but homophony with a word that is a category exemplar.

Note also that the service of the pseudohomophone spelling controls *as* spelling controls is not compromised by their pseudohomophony. It is only their spelling that is important for their control effect upon the categorization response. Their sound is orthogonal to the principle manipulation concerning effects upon false positive categorization responding because, although they are each similar in spelling to some category exemplar, they do not sound like an exemplar of any of the categories that were used in this experiment. For this reason, present concerns were not compromised by the inclusion of this pilot experiment within Experiment 1.

foils (indicated by the means) can in fact be attributed to a small part of the "no" latency distribution. These two distributions of "no" RTs are only different in their long RT tails; otherwise there is virtually no difference between the homophone foils' "no" RT distribution and the spelling control foils' "no" RT distribution.

This observation suggests that were we to excise just the long RT tails of these distributions and then recompute the means, we might not find any difference at all. Table 3 shows repeatedly recomputed mean response times for word homophone foils and word spelling controls as the underlying distributions of correct "no" latencies are repeatedly truncated using successively faster upper cutoff RTs.[2] Notice that the original difference of 87 ms is reduced to a difference of −4 ms with an upper bound of 1,300 ms. Note also that slightly less than 30% of the original data points (the data points used to calculate the original means at the highest upper bound) were excised by an upper bound of 1,300 ms.

The extent of the overlap between the homophone and spelling control "no" RT distributions is underscored by the fact that such a small portion of the original data was truncated, even though we projected our strict yoking of trial outcomes into the truncation procedure. (That is, whenever a trial RT was truncated for a word homophone foil or a spelling control foil, the corresponding, yoked spelling control or homophone trial RT was also always deleted.) Additionally, approximately one third of the excised, yoked trial pairs at the upper bound of 1,300 ms were word-homophone/spelling-control pairs in which the response time to the word homophone was *faster* than the corresponding spelling control response time.

Van Orden (1987) also observed this similarity between the "no" RT distributions of homophone foils and control foils. From this observation, and other data, he hypothesized that subjects use a spelling check (verification procedure) in word identification (at least in experiments with homophonic foils). By his logic, outlier "no" RTs to homophone foils may come from trials in which subjects get "stuck" between the acceptance and rejection criteria used in the spelling check, possibly resulting in further iterations of memory retrieval and verification.

"Yes" latencies. Mean "yes" response times are presented in Table 4. No significant difference was found between false positive "yes" response times to nonword homophone foils

Table 3
Successive Mean Correct "No" Response Times (RTs, in Milliseconds) to Word Homophone Foils Like HARE and Word Spelling Control Foils Like HARP, and Their Differences, as the Underlying RT Distributions Are Repeatedly Truncated Using Successively Faster Cutoff RTs

	Cutoff RT (in ms)				
Foil	2,000	1,400	1,300	1,200	1,100
HARE	979	854	803	783	754
HARP	892	838	807	777	762
Difference	87	16	−4	6	−8

Table 4
False Positive "Yes" Response Times (in Milliseconds) to Nonword and Word Homophone Foils Like SUTE and HARE, Respectively, and Corresponding Correct "Yes" Response Times to Exemplar Controls Like DRESS and TOOTH: Experiment 1

Foil	*M*	*SE*	Control	*M*	*SE*
SUTE	795	43	DRESS	775	38
HARE	825	51	TOOTH	817	35

like SUTE (795 ms) and correct "yes" response times to actual category exemplars like DRESS (775 ms), $t(24) = .49$, for subjects, $t(4) = 1.22$, $p = .29$, for items.[3] Likewise, no significant difference was found between false positive "yes" response times to word homophone foils like HARE (825 ms) and correct "yes" response times to category exemplars like TOOTH (817 ms), $t(25) = .19$, for subjects, and $t(5) = 1.07$, $p = .33$, for items.

This failure to observe a significant difference between false positive and correct "yes" latencies is especially surprising because the statistical test used in these comparisons was relatively liberal. We chose to compare each of a subject's false positive "yes" response times with the mean of the same subject's "yes" response times to six corresponding category exemplars. For example, if a subject categorized SUTE as AN ARTICLE OF CLOTHING, then the time taken for this response was compared with the mean of that same subject's "yes" response times to the category exemplars DRESS, SHOES, PANTS, BLOUSE, SOCKS and VEST. (If a subject incorrectly responded "no" to any of the six control exemplars, then that trial was excluded from the calculation of the corresponding, mean, control, category exemplar, "yes" RT. On average, subjects correctly categorized control exemplars on 93.1% of trials). This method substantially reduces the variance of the distribution of correct "yes" response times. Consequently, if a difference exists between false positive "yes" response times and correct "yes" response times, then we should be more likely to detect it.

These analyses of "yes" latencies provide little evidence that the categorization process (including word identification and meaning evaluation) that precedes a false positive "yes" response to a homophonic foil like SUTE differs from the categorization process that precedes a correct "yes" response to an actual category exemplar like DRESS. Further positive evidence that the two kinds of items are processed similarly comes from a post hoc correlational analysis. In this analysis, "yes" response times to both category exemplars and homo-

[2] We did not illustrate this for the nonword stimuli because, as we noted in Footnote 1, some of their spelling controls were pseudohomophones of words that were not category exemplars, and these pseudohomophone spelling controls seem also to generate exaggerated "no" latencies.

[3] In the item analyses, to compensate for the uneven distribution of errors across items, we formed "super" item means that combine the "yes" response from the homophone foils with error rates less than 20%.

phone foils were found to vary in a similar way across the various categories used ($r = .69$), $t(16) = 3.84$, $p < .05$.

This observed similarity between false positive "yes" response times to homophonic foils and correct "yes" response times to their corresponding category exemplars is hard to reconcile with most current models of word identification. Models that do not include phonological sources of activation are obviously in trouble because they cannot explain the basic phenomenon of false positive responses to homophone foils. But, even models that allow phonological activation to influence word identification almost all presume that phonological activation occurs later than direct orthographic activation (Allport, 1977; Coltheart, 1978; McCusker et al., 1981; Seidenberg, 1985). Thus, there are good theoretical reasons to suspect the validity of this null finding.

It is possible that the analysis of response times in Experiment 1 was too coarse grained to detect the time delay predicted by current, dominant models. Our control correct "yes" response time means are computed from the RTs to six category exemplar items that span a range of typicality that includes the typicality rating of the "sound-alike" category exemplar corresponding to the homophone foil. Suppose that category exemplars that are low in typicality, corresponding to longer correct "yes" RTs, skew the underlying distribution of the control "yes" latencies toward longer RTs. Extra slow "yes" RTs to untypical category exemplars could cause the mean "yes" RT of each category's five control exemplars to appear as slow as the false positive "yes" latencies to homophonic foils, and would thus mask the time delay predicted by most current models of word identification.

Experiment 2

Experiment 1 resulted in two surprising null findings. Word homophone foils like HARE and nonword homophone foils like SUTE produced virtually identical false positive error rates. Also, correct "yes" RTs to category exemplars like DRESS and TOOTH were not significantly faster than false positive "yes" RTs to homophone foils like SUTE and HARE. The purpose of Experiment 2 is to test again for effects of lexicality upon false positive error rates, and to test again for a difference between correct "yes" response times and false positive "yes" response times. To that purpose, Experiment 2 includes a comparison of false positive error rates between new sets of yoked word and nonword homophone foils and a comparison between each false positive "yes" response time (to a homophonic foil) and the same subject's correct "yes" response time to a matched control exemplar of the corresponding category. These control category exemplars are used to allow a more fine-grained comparison between "yes" latencies than was possible in Experiment 1. (The means of selecting these yoked category exemplar targets is described in the method section.)

Method

Subjects. A new group of 50 students was recruited from the subject pool used in Experiment 1.

Procedure. Generally, increasing subjects' error rates should increase their false positive error rates nearer to the range of error rates close to 50%. We assumed that this range would be most sensitive for detecting potential differences in false positive error rates to word and nonword homophones. Accordingly, in Experiment 2 slightly more emphasis was put on responding quickly: After the last practice trial and before the first experimental trial, subjects were reminded that they should continue to try to respond accurately, but *very* quickly. Otherwise, the procedure was identical to that used in Experiment 1.

Stimuli. A new set of stimuli was constructed using criteria similar to those of Experiment 1. The criteria differed in that nonword and word homophone foils were matched for category exemplar frequency (mean log frequency = 1.08, SD = 0.45, for nonword homophones and mean log frequency = 0.92, SD = 0.56, for word homophones) and spelling similarity (OS = .63, SD = .093, for nonword homophones and .58, SD = .068, for word homophones) to their respective category exemplars, only, but not for typicality of corresponding category exemplars (the homophone foils used in Experiment 2 are listed in Appendix C).

Typicality of the homophones' respective category exemplars was not matched because Experiment 1 had, effectively, exhausted the pool of homophones from Uyeda and Mandler's (1980) typicality norms. Fortunately, however, a partial correlational analysis, using the data from Experiment 1, had revealed that the partial correlation between false positive error rates to homophone foils like SUTE and the typicality of corresponding category exemplars like SUIT was not significantly independent of the effect of category exemplar (SUIT) frequency (partial r = .09). But, the partial correlation between category exemplar (SUIT) word frequency and the error rate to homophone foils like SUTE is significantly independent of the typicality effect (partial $r = -.49$), $t(17) = 2.09$, $p = .05$.[4]

Additionally, spelling controls were not included in Experiment 2. The added constraint of needing to find good spelling control items severely restricts the pool of possible stimuli. The difference in error rate between nonword homophones and their spelling controls, found in Experiment 1, is highly significant and the analogous difference in error rate between word homophones and respective spelling controls has been replicated elsewhere (Van Orden, 1987). Therefore, we decided that the costs of providing spelling controls in Experiment 2 outweighed the benefits.

Category exemplar controls for comparison of "yes" response times. Each word and nonword homophone foil (e.g., HARE and SUTE) used in Experiment 2 is homophonic to a particular category exemplar (e.g., HAIR for the category A PART OF THE HUMAN BODY and SUIT for the category AN ARTICLE OF CLOTHING), and in this experiment we planned to compare the false positive "yes" response times to homophone foils like SUTE with the correct "yes" response times to actual category exemplars like SUIT. It seemed to us that the most straightforward way to proceed with this comparison would be to present both SUTE and SUIT to each subject. However, in pilot work we found that when subjects are presented with both SUIT and SUTE in the same experimental session they seem to notice the "trick" and adopt a very conservative categorization strategy, as reflected in decreased error rates and extra long response times. If we are to generate a healthy error rate, then we must avoid presenting subjects with both SUTE and SUIT. Consequently, some other method is required to sample responses from yoked stimuli within the same subject's experimental session.

[4] The exemplars in the typicality norms of Uyeda and Mandler (1980) are those used in the production frequency norms of Battig and Montague (1969). This made it possible to test for an effect of production frequency; the correlation between production frequency of category exemplars (SUIT) and the error rate to homophones (SUTE) was not significant (r = .07).

Although SUIT cannot appear together with its corresponding homophone foil, some other article of clothing could "stand in" for it, say DRESS. If it were established that correct "yes" response times to SUIT and DRESS do not differ, and it were also true that false positive "yes" response times to SUTE do not differ from correct "yes" response times to DRESS, then we can plausibly conclude that false positive "yes" response times to SUTE do not differ from correct "yes" response times to SUIT. This is the logic of the comparison of "yes" response times in Experiment 2.

To choose the "stand-in" control items, we tested 15 high school students in a categorization experiment similar to Experiment 2 except that actual category exemplars like SUIT were substituted for corresponding homophone foils like SUTE. Mean item "yes" latencies for trials in which RTs were greater than 150 ms and less than 1,328 ms (1,328 ms is approximately 3 *SD*s above the mean of the distribution of trial "yes" latencies to targets like SUIT) were compared between items like SUIT and five other exemplars of the appropriate category (e.g., AN ARTICLE OF CLOTHING). The category exemplar that produced the mean response time closest to the mean response time to a target like SUIT was chosen to be the yoked control category exemplar for the corresponding homophone foil (SUTE). These control category exemplars appear in Appendix C with their corresponding homophone foils.

Results and Discussion

Two trials of interest with response times less than 150 ms (presumably anticipations) were excluded from all analyses.

False positive errors. Error rates and "yes" response times to nonword and word homophone foils are shown in Table 5. Although false positive error rates were, as planned, higher than in Experiment 1, the false positive error rate to nonword homophones like SUTE (32.5%) was again virtually equal to that for word homophone foils like HARE (32.8%, $t < 1$, for both subject and item analyses). Thus, the surprising equivalence of false positive errors to nonword and word homophones in Experiment 1 was confirmed in Experiment 2. Because the same result held with two entirely different sets of items, and with a total of 80 subjects in the two experiments together, we take seriously the lack of any substantial difference. The simplest conclusion is that Route 2 (in Figure 1), which involves retrieval of a phonological representation subsequent to lexical access of the stimulus word itself, is virtually never used. However, in the general discussion we will consider some more complex explanations for the apparent equivalence.

Table 5
Percentage of False Positive Categorization Responses and False Positive "Yes" Response Times (RTs) in Experiment 2 to Nonword and Word Homophone Foils Like SUTE *and* HARE*, Respectively, and Corresponding Correct "Yes" Response Times to Yoked Exemplar Controls Like* DRESS *and* TOOTH

% False positives			"Yes" RTs (in ms)					
Foil	*M*	*SE*	Foil	*M*	*SE*	Control	*M*	*SE*
SUTE	32.5	3.3	SUTE	793	43	DRESS	730	38
HARE	32.8	3.0	HARE	797	40	TOOTH	748	35

Table 6
Successive Mean False Positive and Correct "Yes" Response Times (RTs in Milliseconds) to Nonword and Word Homophone Foils Like SUTE *and* HARE*, and to Corresponding Yoked Exemplar Controls Like* DRESS *and* TOOTH*, and Their Respective Differences, as the Underlying RT Distributions Are Repeatedly Truncated Using Successively Faster Cutoff RTs*

	Cutoff RT				
Stimulus	2,000	1,500	1,000	900	800
SUTE	793	764	642	605	576
DRESS	730	691	610	586	576
Difference	63	73	32	19	0
HARE	797	772	637	589	574
TOOTH	748	705	637	612	588
Difference	49	67	0	−23	−14

"Yes" latencies. Trials with response times under 150 ms or greater than 2 s were again dropped from the analysis of response times. The mean of false positive "yes" response times to nonword homophonic foils like SUTE (793 ms) was slower than the mean of the correct "yes" response times to matched actual category exemplars like DRESS (730 ms). However, this difference was only significant in the analysis across subjects: $t(44) = 2.00$, $p = .05$, for subjects; $t(9) = 1.69$, $p > .12$, for items. The mean of the incorrect "yes" response times to word homophone foils like HARE (797 ms) was also slightly slower than the mean of the correct "yes" response times to matched actual category exemplars like TOOTH (748 ms). However, this difference was not statistically reliable in either analysis: $t(45) = 1.21$, $p > .20$, for subjects; $t(9) = 1.29$, $p > .20$, for items.

Inspection of these distributions revealed that the small differences between the mean "yes" RTs to category exemplars and homophonic foils were not due to a shift in the overall distributions. Rather, just as we found for "no" latencies in Experiment 1, the difference in the distributions was concentrated at their high (slow RT) tails, where there were more outlier RTs for homophone foils (both word and nonword) than in the corresponding portion of the correct "yes" RT distribution. Thus, once again, if we were to truncate these distributions just before the outlier RTs, and then recompute the means, the difference should disappear. This point is illustrated in Table 6, which gives the mean times for the relevant conditions as the underlying distributions are repeatedly truncated using successively decreasing upper cutoff RTs. (In order not to bias comparisons due to differences in item properties, whenever a trial was lost for a homophone item, the corresponding yoked correct "yes" trial was deleted, and vice versa.) The means approximately converge for the SUTE versus DRESS distributions with an upper cutoff of 800 ms, and for HARE versus TOOTH distributions with an upper cutoff of 1,000 ms.

Note that slightly more than 30% of the original word homophone and exemplar control data points (the data points used to calculate the original means at the highest upper bound) were excised by an upper bound of 1,000 ms, and that slightly more than 40% of the original nonword homophone and exemplar control data points were excised by an upper bound of 800 ms. Once again, just as for "no" RTs in Experiment 1, the extent of the overlap between the incorrect "yes" RTs to homophone foils and correct "yes" RTs to exemplar controls is underscored by the fact that such small portions of the original data were truncated, even though we projected our strict yoking of trial outcomes into the truncation procedure. Additionally, of the word-homophone/exemplar trial pairs that were excised at the upper bound of 1,000 ms, approximately 38% were trial pairs in which the incorrect "yes" response time to the word homophone was faster than the correct "yes" response time to the corresponding yoked exemplar. Similarly, of the nonword-homophone/exemplar trial pairs that were excised at the upper bound of 800 ms, approximately 28% were trial pairs in which the incorrect "yes" response time to the nonword homophone was faster than the correct "yes" response time to the corresponding yoked exemplar.

It seems fair to summarize the data as showing that false positive "yes" responses to word and nonword homophones are, generally, little if any slower than the corresponding correct "yes" responses to actual category exemplars. This correspondence is especially close for the fast tail of the distributions. However, before speculating about the import of this close correspondence, a caution is necessary. We are not comparing two distributions of the same kind. Rather, we are comparing a randomly sampled selection of correct "yes" trials, to a subject-selected sample of false positive trials that is unlikely to be randomly sampled with respect to many variables that would affect response time.

For example, false positive errors might tend to come from trials in which subjects' decision criterion fluctuated in a more lax direction. If so, then we cannot know what the data would look like for the corresponding subset of correct "yes" trials equivalent in criterion (or any other hypothetical source of a selection artifact). If, instead of our selection of correct "yes" trials based on item yoking, we could magically look at the selection of correct "yes" trials with the same lax matching criteria, that distribution might be shifted toward faster responses.

Counter to this artifact hypothesis is the fact that the fastest "yes" response times to control category exemplar targets (irrespective of item yoking), response times that estimate the lower bound for those correct "yes" trials that result from a lax decision criterion (or some other selection artifact), are no faster than the fastest false positive "yes" response times. Also, it is at least consistent with a no-artifact hypothesis that a similar pattern of overlap was observed between homophone and control foil "no" RT distributions in Experiment 1; data for which a selection artifact is extremely unlikely. That is, homophone foils and their "yes" response controls in Experiment 2 produced the same pattern of RT distribution overlap as homophone foils and their "no" response controls in Experiment 1.

In any case, with this caution noted, we will now consider some possible implications of the near equivalence of correct "yes" response times and false positive "yes" response times to homophone foils. Most straightforwardly and conservatively, we can conclude that phonological mediation does indeed occur rapidly enough to influence responses in the same range of observed times as "yes" responses, answering the criticism of DYME'S effect in the lexical decision task that we noted in the introduction to Experiment 1. The present result is especially powerful in this respect, because the final response pathway of correct and false positive "yes" responses is hypothesized to be the same. Thus, there is no good reason to doubt that phonological mediation influenced the key internal event that we are concerned with—word identification—during its normal time course.

The point just made would hold merely from the observation of a large overlap in the response time distributions. But, if, in spite of the foregoing caveat, we could conclude that homophone foils actually produced false positive "yes" responses as rapidly as correct "yes" responses, much stronger conclusions might be warranted. Note that homophone foils are not spelled exactly like category exemplars, but, of course, category exemplars are spelled exactly like themselves. This surely means that a direct spelling to lexical node connection should provide much stronger activation for a true category exemplar than for a homophonic foil. In contrast, if we knew that the false word identification caused by homophone foils occurred as rapidly as veridical word identification of true category exemplars, then we would appear to have evidence that the direct spelling to lexical node route plays no substantial role at all. This would be an astonishing conclusion; almost no recently presented models would predict such a result.

Perhaps the most promising way of explaining such a result would be a very extreme version of a verification model (e.g., see Van Orden, 1987) in which phonological mediation provides exclusive bottom-up activation of lexical candidates, and spelling is checked in a verification procedure. In our "yes" RT data the only effect of spelling dissimilarity is the extended tail at the high end of the distribution of incorrect "yes" RTs to homophonic foils. Current dual access models would predict a general effect of orthographic dissimilarity, not an occasional increase in response times. But, in a verification model, these exaggerated RTs could reflect trials in which the verification procedure got "stuck" between acceptance and rejection criteria, resulting in additional cycles of memory retrieval and comparison.

A final point concerns whether subjects knew the correct spellings of the category exemplars that are homophonic to our stimulus homophone foils. In auxiliary experiments, we tested high school students' knowledge of the meanings of the category exemplars. Subjects were presented with each of the category exemplars (written on a blackboard but not named) and then asked to write a sentence that clearly showed what the word meant (e.g., an unacceptable response would be "HAIR is a word," but an acceptable response would be "The HAIR on his head is brown"; an example like this was used in the instructions to subjects). All subjects responded with sentences that clearly demonstrated their knowledge of the meanings that are associated with the spellings of exemplars like HAIR and SUIT. This was true for all of the exemplars in both Experiments 1 and 2. (Two exemplars, PLANE and ROSE, from Experiment 1 were not usually used in subjects' sentences to indicate a "carpenter's tool" or "a color," respectively, but every other exemplar was virtually always used in a sentence whose meaning was consistent with its "category meaning.")

Van Orden (1987) reported more direct tests of subjects' spelling knowledge for those exemplars (HAIR) that corresponded to his stimulus homophone foils (HARE). Subjects in his spelling tests showed near perfect knowledge of category exemplar spellings (except for a very few spelling errors to very low frequency category exemplars). He interpreted this discrepancy between spelling test results (almost no misspellings of category exemplars) and categorization results (high false positive error rates to homophone foils that, of course, are not spelled exactly like category exemplars) to mean that complete knowledge of category exemplar spelling alone (as demonstrated by spelling performance) does not prevent subjects from making false positive errors to homophone foils. Rather, in addition to being relatively complete, this spelling knowledge must be readily available to the verification procedure.

General Discussion

The present experiments replicate Van Orden's (1987) finding that, in a categorization task, target words that are homophonic to category exemplars (e.g., HARE as A PART OF THE HUMAN BODY) produce far more false positive categorizations than control words that are equally close in spelling but not homophonic to category exemplars. Two further results were observed in error data: (a) Homophonic nonwords (e.g., SUTE as AN ARTICLE OF CLOTHING) also produced far more false positive errors than spelling controls, and (b) false positive errors occurred equally often for nonword homophones and word homophones (in both Experiments 1 and 2).

Computed Phonology and Reading

Of these two new results, the first is the most important because it provides direct positive evidence for the use of computed phonological representations in word identification. Van Orden's (1987) previous result, using word homophones, like HARE, might have been caused by a phonological representation retrieved at a lexical entry corresponding to the target word (e.g., "HARE") that mediated access to a category exemplar like HAIR (see Route 2 in Figure 1). Such a roundabout route cannot be present for foils like SUTE because nonwords have no lexical entries; so we are left with the classical phonological mediation pathway (Route 1 in Figure 1) as the explanation of the elevation in false positive errors.

Equivalence of Word and Nonword Homophone Errors

The second finding from our error results is that nonword homophone foils like SUTE and word homophones like HARE produce equally high false positive error rates. The simplest interpretation of this result is that retrieved phonological representations (Route 2 in Figure 1), which could only be available for word homophones, play no role in generating errors. According to this interpretation, the route that uses computed phonological representations (Route 1) is the only source of errors to word homophone foils as well as nonword homophone foils. The argument that nonword and word homophones are processed similarly gains support from the remarkably close response time distributions for false positives in Experiment 2 (M = 793 ms, SE = 43, for nonword homophones; M = 797 ms, SE = 40, for word homophones). Note that the hypothesis that Route 2 plays no role in our data need not mean that this route can never be used; it may be too much *slower* than Route 1 to influence performance on our task (contrary to the assumptions of Allport, 1977; Coltheart, 1978; McCusker et al., 1981).

Although we will argue that the "Route 1 only" hypothesis is the most attractive one, there are more complex possibilities. In particular, the underlying tendency for Route 2 to produce more errors for word homophones could have been masked by additional processing paths that counter that tendency. As it happens, some additional pathways have considerable a priori plausibility. Figure 2 shows a processing diagram like Figure 1, but with some additional paths that are candidates for affecting processing of word homophones.

Consider first Pathway 3. Activation of the lexical entry "HARE" might plausibly activate its corresponding semantic representation. To the extent that this representation influences the decision process, it will presumably reduce the

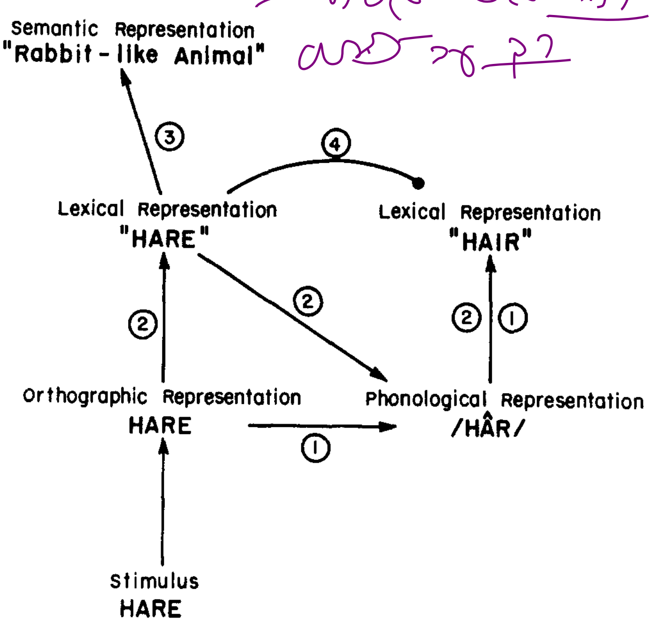

Figure 2. Phonological sources of activation at lexical entry "HAIR" when HARE is the categorization target (from Figure 1) with the added prophylactic Pathways 3 and 4. (Pathways 3 and 4 [readily available semantic representation and inhibitory competition between lexical entries] could equilibrate the effect of Route 2 [the postlexical phonological representation] upon false positive error rates to word homophone foils.)

tendency to accept the target HARE as a A PART OF THE HUMAN BODY. Figure 2 also shows a possible inhibitory connection between the lexical entries "HAIR" and "HARE" (Pathway 4), which would also tend to reduce false positive errors. Inhibition between competing lexical entries has been hypothesized in recent activation models of word identification (e.g., see McClelland & Rumelhart, 1981; Rumelhart & McClelland, 1982). Note that neither Pathway 3 nor Pathway 4 would have any effect on false positive errors to a nonword homophone like SUTE, because no lexical entry exists for this target string.

The possibility that Pathways 3 and 4 (see Figure 2) offset the effect of Route 2 is difficult to rule out entirely; however, this hypothesis is made less likely by the following auxiliary analyses of the pattern of errors in Experiments 1 and 2 and corresponding data from Van Orden (1987). These auxiliary analyses follow from the assumption that the effectiveness of Pathways 3 and 4 depends on how readily the lexical representation for the target word (e.g., "HARE") can be activated. Relatively familiar target words (as indexed by frequency) should activate their lexical representations more strongly/rapidly. Thus, we should expect Routes 3 and 4 to have a greater effect for higher frequency words, an effect that should be evident in fewer errors to higher frequency target words.

Contrary to this hypothesis, an analysis of the combined data from Experiments 1 and 2 showed virtually no correlation between the frequency of the target foil itself (e.g., HARE) and error rates ($r = -.08$). Van Orden (1987) also failed to find any effect of word homophone foil frequency upon false positive error rates in his Experiment 3 which was designed to test for that possibility. Null results must be interpreted cautiously, but a similar analysis was sensitive to the effects of the frequency of the corresponding category exemplar (e.g., HAIR) on false positive error rates to corresponding homophone foils (e.g., HARE), both for the present data, $r = -.48$, $t(18) = -2.32$, $p < .05$, in Experiment 1, and $r = -.42$, $t(18) = -1.97$, $p = .06$, in Experiment 2 and in Van Orden's (1987) Experiment 3.

Please note that the foregoing failure to find an effect of stimulus HARE's frequency does not allow the inference that stimulus meaning activation plays no role in categorization. Rather, we may only infer that the stimulus homophone HARE's activation of HAIR's meaning, presumably the source of false positive categorization errors, is seemingly unqualified by any possible parallel or prior activation of a "RABBIT-LIKE ANIMAL."

Persistent supporters of Pathways 3 and 4 might wish to argue that the above argument concerning word homophone frequency is itself open to the possibility that the true effects of frequency on these paths are actually masked by the effects of frequency on Routes 1 and 2. If target foil familiarity increased the potency of Routes 1 and 2, then the resulting increase in errors could offset the expected decrease in errors from Pathways 3 and 4.

It may appear that we are caught in a revolving door at this point, but an exit is possible. If we could eliminate the effects of Routes 1 and 2, then the efficacy of Paths 3 and 4 could be tested cleanly. Fortunately, the word spelling controls provide a way to perform this test. Figure 3 shows a processing diagram for the target HARP that serves as a spelling control for the word homophone target HARE. HARP's phonological representation (whether activated via Route 1, or via Route 2) should not appreciably activate the lexical representation "HAIR" because HARP, unlike HARE, is not a homophone of HAIR. The lexical representation of HARP, in contrast, has the same possibilities of reducing false positive errors via Paths 3 and 4 as did the lexical representation for HARE. Thus, we can carry forward the earlier logic, and predict that (if Paths 3 and 4 are effective) false positive errors should be lower for spelling control targets that are higher in frequency. In other words, the effect of Paths 3 and 4 would be revealed in a significant negative correlation between the frequency of spelling control words and their false positives error rates.

In fact, we obtained a negligible correlation in the wrong direction ($r = .08$) for the data in Experiment 1. This result was corroborated by an even smaller correlation ($r = .03$), again in the wrong direction, obtained from a reanalysis of similar spelling control data obtained by Van Orden (1987). We also obtained a slightly larger correlation ($r = .22$), in the wrong direction in a third test, reanalyzing spelling control data from another categorization experiment of Van Orden (1987), where the error rate was increased by pattern-masking targets. Thus, we conclude that, in spite of the a priori plausibility of Paths 3 and 4, there is no evidence that they play any measurable role in the categorization task. If Paths 3 and 4 can be discounted, we are left with the simple but elegant explanation for the equal error rates to word and nonword homophones with which we began. All homophones, whether words or nonwords, produce false positive categorization errors through a computed phonological rep-

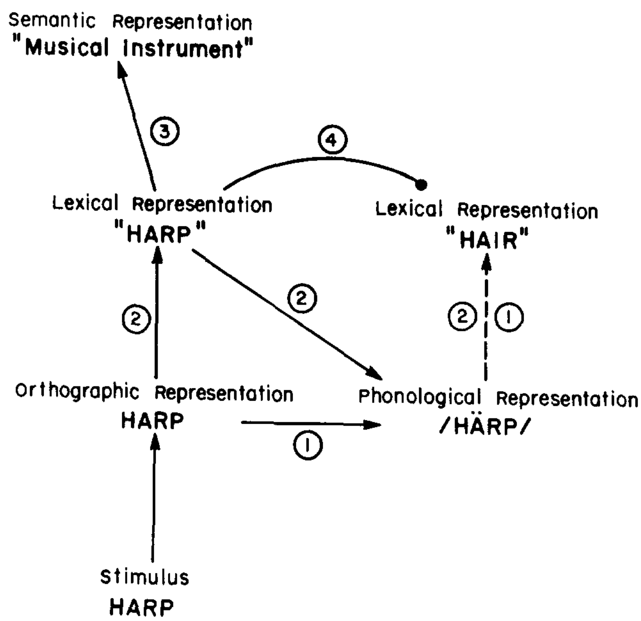

Figure 3. The implications for Routes 1 and 2 relative to Pathways 3 and 4 when the target is a spelling control like HARP. (Although Routes 1 and 2 are no longer potent sources of false positive categorization errors, Pathways 3 and 4 remain available to avoid categorization errors.)

resentation (Route 1), not a retrieved representation (Route 2).

Independent Direct Access?

The dual-process model (Coltheart, 1978), possibly the most influential of the dual access models, assumes that for skilled readers the primary difference between words and nonwords is the possibility of direct lexical access for words, but not for nonwords. This assumption would be supported if we had found any difference in categorization performance to word and nonword stimuli. However, we have failed to observe any influence of the independent direct access route under seemingly optimal conditions for this observation.

Direct access is assumed to be the source of readers' ability to understand correctly homophonic words (Doctor & Coltheart, 1980; Ellis, 1984; but cf. Van Orden, 1987); thus, measures of performance to homophonic stimuli should be especially sensitive to the effects of the direct access route. Furthermore, the availability of direct access is presumed to be directly related to stimulus familiarity (usually estimated by frequency), and although our method has proved to be sensitive to the effects of category exemplar frequency (an effect predicted by a verification alternative to dual process theory; see Van Orden, 1987), no effect of homophone stimulus frequency was observed in any of the present experiments or in the experiments reported by Van Orden (1987).

Given the direct access route's reluctance to show itself in our data, it might pay to reexamine some of the empirical findings and a priori assumptions that motivate this hypothesis. Van Orden, Pennington, and Stone (1988) reviewed this literature and concluded that no unambiguous support exists for an independent route of direct access. Van Orden et al.'s review, coupled with the auxiliary analyses of our own data, at the very least, suggest that the intuitively appealing, putative, "independent direct access hypothesis" may not rest on strong empirical support.

Phonological Coding of Irregular Words

Consider BREAK, a word homophone foil from Experiment 2 corresponding to the category PART OF A BICYCLE. BREAK's spelling to sound correspondence is irregular (e.g., see Coltheart, Besner, Jonasson, & Davelaar, 1979). The existence of irregular words like BREAK has always been the cornerstone of the independent direct access hypothesis. This traditional assumption of dual process theory presupposes that irregular words require direct access (Route 2) because they are exceptions to the grapheme-phoneme correspondence rules that govern Route 1. Thus, our subjects' error rate of 54% to the word homophone foil BREAK would seem to be evidence for an effect of retrieved phonological representations.

The dual process analysis of regularity has, however, been critically undermined by studies that show consistency to be the predominant source of regularity effects (Andrews, 1982; Bauer & Stanovich, 1980; Glushko, 1979; 1981). Also, in contradiction to the "irregular words require direct access assumption," Van Orden (1987) has proposed a connectionist mechanism of phonological coding that can learn to code both irregular and regular words. Sejnowski and Rosenberg (1986; see also Rosenberg & Sejnowski, 1986) used a related mechanism in artificial intelligence (AI) simulations of reading aloud. Their model clearly demonstrated a connectionist model's ability to phonologically code irregular words.

This is not to say that we deny the possibility of direct bottom-up activation of lexical features by orthographic features. Rather, it may merely be useful to abandon the notion of separate, independent routes of lexical access. A potential alternative to the independent routes hypothesis is a connectionist mechanism (like those noted earlier) that, through learning, comes to reflect the covariance between all linguistic features (syntactic, semantic, and phonological) and orthographic features in its associative weights (Van Orden, 1987). This matrix of associative weights could then function to transform orthographic representations into lexical representations. This view assumes that lexical codes are composed of morphophonological features (see also Chomsky, 1970; Fowler, Napps, & Feldman, 1985; Lima & Pollatsek, 1983; Murrell & Morton, 1974; Snodgrass & Jarvella, 1972; Taft, 1979, 1981; Taft & Forster, 1975), and that orthographic features activate all lexical features—whatever their nature—via a common connectionist mechanism.

Reading in the Categorization Task Versus Reading in General

It is always an open question whether experimental reading tasks generalize to more normal reading situations. Consequently, it is useful in this regard to examine the observed phonological effects in terms of the processing requirements of the categorization task.

The categorization task, like the reading of text, requires that a target word be understood relative to the immediate context in which it appears. Because both categorization and text understanding require access to a word's meaning, we assume that they both require word identification. Thus, in this regard, reading in the categorization task compares well with normal reading, much better than the lexical decision judgment of wordness.

The categorization task also allowed us to test for effects of phonology using both real words and nonword targets. We have already noted the theoretical implications of effects of nonword phonology. But, it may be more important for inferences concerning normal reading that we (and Van Orden, 1987) observed effects of real word phonology. After all, everyday reading is primarily the reading of relatively common real words.

Additionally, because the present categorization tasks included only a small percentage of homophone foil trials (10% of experimental trials), we minimized the possibility of inducing processing strategies that are observed when a large proportion of trials contain homophone foils (e.g., see Davelaar et al., 1978; Hawkins, Reicher, Rogers, & Peterson, 1976; McQuade, 1981).

This is not to say, however, that reading in the categorization task is identical in all respects to the reading of text. The evaluation within the categorization task of a target "X" appears to us to be similar to the evaluation of the truth of the proposition "X is a Y." We do not imagine that this is the primary method by which individual words in text are understood relative to their context. But, we do assume that the process of word identification operates similarly in both cases, supplying information from memory about a particular word's meaning. Thus, we believe that performance in the categorization task can (and in the present case does) reveal the mechanisms of normal reading, especially the mechanisms of word identification. Consequently, we infer that computation of phonological codes for the process of word identification is a necessary, if not fundamental, component of reading.

Summary and Conclusions

A primary finding of Experiment 1 is that nonword homophones are mistaken for sound-alike words in a categorization task, a result that can be attributed to their phonological identity. Thus, computed phonological codes are a potent force in word identification. A secondary finding of Experiment 1, replicated in Experiment 2, is that matched nonword and word homophone foils are nearly equally likely to be mistaken for sound-alike category exemplars. This finding is consistent with a theory of word identification in which the mechanism of lexical coding is blind to stimulus familiarity, at least when unfamiliar stimuli mimic exactly the phonology of actual words. Thus, considering only this data, there seems little danger of overestimating the role of phonology in lexical coding. Most current theories of lexical coding do not include a role of importance for phonology, if they include any role at all. This is easily understood because, until recently, the case for phonological mediation has lacked clear supporting evidence. But, if the present analysis is correct, then that evidence now exists.

References

Aaronson, D., & Ferres, S. (1983). A model for coding lexical categories during reading. *Journal of Experimental Psychology: Human Perception and Performance, 9*, 700–725.

Allport, D. A. (1977). On knowing the meaning of words that we are unable to report: The effects of visual masking. In S. Dornic (Ed.), *Attention and performance VI* (pp. 505–533). New York: Academic Press.

Andrews, S. (1982). Phonological recoding: Is the regularity effect consistent? *Memory & Cognition, 10*, 565–575.

Baddeley, A. D. (1979). Working memory and reading. In P. A. Kolers, M. E. Wrolstad, & H. Bouma (Eds.), *Processing of visible language* (Vol. I, pp. 355–370). New York: Plenum Press.

Baddeley, A. D., Eldridge, M., & Lewis, V. (1981). The role of subvocalization in reading. *Quarterly Journal of Experimental Psychology, 33A*, 439–454.

Baddeley, A. D., & Lewis, V. (1981). Inner active processes in reading: The inner voice, the inner ear and the inner eye. In A. M. Lesgold & C. A. Perfetti (Eds.), *Interactive processes in reading* (pp. 107–129). Hillsdale, NJ: Erlbaum.

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

Barron, R. W. (1979). Reading skill and phonological coding in lexical access. In M. M. Gruneberg, R. M. Sykes, & P. E. Morris (Eds.), *Practical aspects of memory* (pp. 468–475). London: Academic Press.

Barron, R. W. (1980). Visual-orthographic and phonological strategies in reading and spelling. In U. Frith (Ed.), *Cognitive processes in spelling* (pp. 195–213). London: Academic Press.

Battig, W. F., & Montague, W. E. (1969). Category norms for verbal items in 56 categories. *Journal of Experimental Psychology Monograph, 80*, 1–46.

Bauer, D., & Stanovich, K. (1980). Lexical access and the spelling-to-sound regularity effect. *Memory & Cognition, 8*, 424–432.

Becker, C. A. (1976). Allocation of attention during visual word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 2*, 556–566.

Becker, C. A. (1980). Semantic context effects in visual word recognition: An analysis of semantic strategies. *Memory and Cognition, 8*, 493–512.

Besner, D., Davies, J., & Daniels, S. (1981). Reading for meaning: The effects of concurrent articulation. *Quarterly Journal of Experimental Psychology, 33A*, 415–437.

Bower, T. G. R. (1970). Reading by eye. In H. Levin & J. P. Williams (Eds.), *Basic studies on reading* (pp. 134–146). New York: Basic Books.

Chomsky, C. (1970). Reading, writing, and phonology. *Harvard Educational Review, 40*, 287–309.

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

Coltheart, M., Besner, D., Jonasson, J. T., & Davelaar, E. (1979). Phonological encoding in the lexical decision task. *Quarterly Journal of Experimental Psychology, 31*, 489–505.

Coltheart, M., Davelaar, E., Jonasson, J. T., & Besner, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.), *Attention and performance VI* (pp. 535–555). New York: Academic Press.

Davelaar, E., Coltheart, M., Besner, D., & Jonasson, J. T. (1978). Phonological recoding and lexical access. *Memory and Cognition, 6*, 391–402.

Doctor, E. A., & Coltheart, M. (1980). Children's use of phonological encoding when reading for meaning. *Memory & Cognition, 8*, 195–209.

Ellis, A. W. (1984). *Reading, writing and dyslexia: A cognitive analysis.* Hillsdale, NJ: Erlbaum.

Fowler, C. A., Napps, S. E., & Feldman, L. (1985). Relations among regular and irregular morphologically related words in the lexicon as revealed by repetition priming. *Memory & Cognition, 13*, 241–255.

Glushko, R. (1979). The organization and activation of orthographic knowledge in reading aloud. *Journal of Experimental Psychology: Human Perception and Performance, 2*, 361–379.

Glushko, R. (1981). Principles for pronouncing print: The psychology of phonography. In A. M. Lesgold & C. A. Perfetti (Eds.), *Interactive processes in reading* (pp. 61–84). Hillsdale, NJ: Erlbaum.

Gough, P. B. (1972). One second of reading. In J. K. Kavanaugh & I. G. Mattingly (Eds.), *Language by ear and by eye* (pp. 331–358). Cambridge, MA: MIT Press.

Gough, P. B., & Cosky, M. J. (1977). One second of reading again. In N. J. Castellan, D. B. Pisoni, & G. R. Potts (Eds.), *Cognitive theory 2* (pp. 271–288). Hillsdale, NJ: Erlbaum.

Hawkins, H. L., Reicher, G. M., Rogers, M., & Peterson, L. (1976). Flexible coding in word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 2*, 380–385.

Henderson, L. (1982). *Orthography and word recognition in reading.* New York: Academic Press.

Kleiman, G. M. (1975). Speech recoding in reading. *Journal of Verbal Learning and Verbal Behavior, 14*, 323–339.

Kolers, P. A. (1970). Three stages of reading. In H. Levin & J. P.

Williams (Eds.), *Basic studies on reading* (pp. 90–118). New York: Basic Books.

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

Levy, B. A. (1978). Speech processing during reading. In A. M. Lesgold, J. W. Pelligrino, S. D. Fokkema, & R. Glaser (Eds.), *Cognitive psychology and instruction* (pp. 123–151). New York: Plenum Press.

Lima, S. D., & Pollatsek, A. (1983). Lexical access via an orthographic code? The basic orthographic syllabic structure (BOSS) reconsidered. *Journal of Verbal Learning and Verbal Behavior, 22,* 310–332.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review, 88,* 375–407.

McCusker, L. X., Hillinger, M. L., & Bias, R. G. (1981). Phonological recoding and reading. *Psychological Bulletin, 89,* 217–245.

McQuade, D. V. (1981). Variable reliance on phonological information in visual word recognition. *Language and Speech, 24,* 99–109.

Meyer, D. E., & Gutschera, K. D. (1975). *Orthographic versus phonemic processing of printed words.* Paper presented at the meeting of the Psychonomic Society, Denver, CO.

Meyer, D. E., & Ruddy, M. G. (1973). *Lexical-memory retrieval based on graphemic and phonemic representation of printed words.* Paper presented at the meeting of the Psychonomic Society, St. Louis, MO.

Murrell, G. A., & Morton, J. (1974). Word recognition and morphemic structure. *Journal of Experimental Psychology, 102,* 963–968.

Naish, P. (1980). Phonological recoding and the Stroop effect. *British Journal of Psychology, 71,* 395–400.

Paap, K. R., Newsome, S. L., McDonald, J. E., & Schvaneveldt, R. W. (1982). An activation-verification model for letter and word recognition: The word-superiority effect, *Psychological Review, 89,* 573–594.

Perfetti, C. A. (1985). *Reading ability.* New York: Oxford University Press.

Rosenberg, C. R., & Sejnowski, T. J. (1986). *The effects of distributed vs. massed practice on NETtalk, a massively-parallel network that learns to read aloud.* Unpublished manuscript.

Rubenstein, H., Lewis, S. S., & Rubenstein, M. A. (1971). Evidence for phonemic recoding in visual word recognition. *Journal of Verbal Learning and Verbal Behavior, 10,* 645–657.

Rumelhart, D. E., & McClelland, J. L. (1982). An interactive activation model of context effects in letter perception: Part 2. The contextual enhancement effect and some tests and extensions of the model. *Psychological Review,, 89,* 60–94.

Saffran, E. M., & Marin, O. S. M. (1975). Immediate memory for word lists and sentences in a patient with deficient auditory short term memory. *Brain and Language, 2,* 420–433.

Seidenberg, M. S. (1985). The time course of phonological code activation in two writing systems. *Cognition, 19,* 1–30.

Sejnowski, T. J., & Rosenberg, C. R. (1986). *NETtalk: A parallel network that learns to read aloud* (Electrical Engineering and Computer Science Tech. Rep. No. JHU/EECS-86/01). Baltimore, MD: Johns Hopkins University.

Smith, F. (1971). *Understanding reading: A psycholinguistic analysis of reading and learning to read.* New York: Holt, Rinehart & Winston.

Snodgrass, J. G., & Jarvella, R. J. (1972). Some linguistic determinants of word classification time. *Psychonomic Science, 27,* 220–222.

Spoehr, K., & Smith, E. (1973). The role of syllables in perceptual processing. *Cognitive Psychology, 5,* 71–89.

Stanovich, K. E., & Bauer, D. W. (1978). Experiments on the spelling-to-sound regularity effect in word recognition. *Memory & Cognition, 6,* 410–415.

Taft, M. (1979). Recognition of affixed words and the word frequency effect. *Memory & Cognition, 7,* 263–272.

Taft, M. (1981). Prefix stripping revisited. *Journal of Verbal Learning and Verbal Behavior, 20,* 289–297.

Taft, M. (1982). An alternative to grapheme-phoneme conversion rules? *Memory & Cognition, 10,* 465–474.

Taft, M., & Forster, K. I. (1975). Lexical storage and retrieval of prefixed words. *Journal of Verbal Learning and Verbal Behavior, 14,* 638–647.

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 thesis, University of California, 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., Pennington, B. F., & Stone, G. O. (1988). *Word identification in reading and the promise of subsymbolic psycholinguistics.* Unpublished manuscript.

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

Winer, B. J. (1971). *Statistical principles in experimental design.* New York: McGraw-Hill.

Appendix A

Nonword and Word Homophone Foils and Their Respective Yoked Spelling Controls: Experiment 1

Category	Homophone	Spelling control
	Nonwords	
A VEGETABLE	KARRET	PARRIT
A FOUR-FOOTED ANIMAL	SHEAP	SHELP
A TREE	OKE	ONK
A PART OF THE HUMAN BODY	BRANE	BLAIN
AN ARTICLE OF CLOTHING	SHURT	SHART
A WEATHER PHENOMENON	HEET	HERT
A VEHICLE	JEAP	JELP
AN ARTICLE OF CLOTHING	SUTE	SURT
A WEATHER PHENOMENON	SLEAT	SCEET
A KITCHEN UTENSIL	BOLE	BOLB
	Words	
A PART OF A BUILDING	SELLER	TELLER
A PART OF A BUILDING	STARE	START
A NON-ALCOHOLIC BEVERAGE	TEE	TEN
A CARPENTER'S TOOL	PLAIN	PLACE
A METAL	STEAL	STEEP
A FOUR-FOOTED ANIMAL	BARE	BEER
A FOUR-FOOTED ANIMAL	DEAR	DYER
A PART OF THE HUMAN BODY	HARE	HARP
A VEGETABLE	BEATS	BELTS
A COLOR	ROWS	ROBS

Appendix B

Experiment 1 required a control for the similarity in spelling that is common between pairs of homophones. Van Orden (1984, 1987) adapted Weber's (1970) estimate of orthographic similarity (GS) for exactly this purpose. Van Orden's estimate (OS) is computed as follows:

OS = (GS of homophone pair)/(GS of category homophone and itself);
$GS = 10\,\{[(50F + 30\,V + 10C)/A] + 5T + 27B + 18E\}$.

F = number of pairs of adjacent letters in the same order shared by word pairs:

HOUSE / HORSE $F = 2$;
EVERY / VERY $F = 3$.

V = number of pairs of adjacent letters in reverse order shared by word pairs:

WAS / SAW $V = 2$.

C = number of single letters shared by word pairs:

SPOT / PUFF $C = 1$;
FAMILY / FUNNY $C = 2$.

A = average number of letters in the two words:

EVERY / VERY $A = 4.5$.

T = ratio of number of letters in the shorter word to the number in the longer:

EVERY / VERY $T = 4/5$.

$B = 1$ if the first letter in the two words is the same; otherwise, $B = 0$.
$E = 1$ if the last letter in the two words is the same; otherwise, $E = 0$.
For example, OS between SUTE and SUIT is

$$\begin{aligned} OS &= (\text{GS of SUTE–SUIT})/(\text{GS of SUIT–SUIT}) \\ &= 10[[\{[50(1) + 30(0) + 10(3)]/4\} + 5\{1\} + 27\{1\} + 18\{0\}]]/ \\ &\quad 10[[\{[50(3) + 30(0) + 10(4)]/4\} + 5\{1\} + 27\{1\} + 18\{0\}]] \\ &= 520/975 = .53. \end{aligned}$$

(Appendix C follows on next page)

Appendix C

Nonword and Word Homophone Foils and Their Respective Exemplar Controls: Experiment 2

Category	Homophone	Exemplar control
	Nonwords	
A TYPE OF GRAIN	WHEET	CORN
A BIRD	CROE	PIGEON
A KIND OF MEAT	BEAF	VEAL
CARPENTER'S GEAR	NALE	SCREW
A WEATHER PHENOMENON	HAYLE	LIGHTNING
PART OF A SHIP	KEAL	RUDDER
A BODY OF WATER	CREKE	STREAM
A PLACE OF CONFINEMENT	JALE	CELL
THINGS IN A WOMAN'S PURSE	KEE	COMB
ANIMAL SOUNDS	RORE	BARK
	Words	
A PART OF A BICYCLE	BREAK	TIRE
AN INSECT	FLEE	MITE
PART OF A PERSON'S FOOT	HEAL	ARCH
BEACH GEAR	PALE	BLANKET
PART OF A LION'S BODY	PAUSE	FUR
FISHING GEAR	REAL	POLE
A KIND OF MEAT	STAKE	VEAL
PART OF A LION'S BODY	TALE	FUR
A BODY OF WATER	SEE	LAKE
PART OF A SHIP	SALE	MAST

Received August 18, 1986
Revision received May 11, 1987
Accepted April 7, 1987 ■