Journal of Experimental Psychology: General
1991, Vol. 120, No. 4, 358-394

0096-3445/91/$3.00

Does Word Identification Proceed From Spelling to Sound to Meaning?

Debra Jared
McGill University
Montreal, Quebec, Canada

Mark S. Seidenberg
University of Southern California

Six experiments addressed the role of phonological information in visual word recognition using a semantic-decision task. Experiment 1 replicated Van Orden's (1987) finding that Ss make more false-positive errors on homophone foils than on spelling controls, indicating phonological activation of meaning. Experiment 2 showed that only lower frequency words yield this effect when broader categories are used. In Experiments 3 and 4, the homophony effect for lower frequency words remained, even though the stimuli included a large proportion of homophones, suggesting that activation of phonological information cannot be strategically inhibited. Experiments 5 and 6 examined effects of homophony on targets that were correct category exemplars and yielded similar results. These studies indicate that in skilled readers, phonological information contributes to the activation of word meaning only for low-frequency words.

The world's orthographies represent different solutions to the problem of representing spoken language in written form. Orthographies differ in the extent to which the written symbols encode phonological information. Historically, orthographies have evolved toward more direct representations of phonology (Henderson, 1982; Hung & Tzeng, 1981). The reason for this trend is unclear, although one explanation may be that phonological information serves a useful function in skilled reading or in learning to read (Rozin & Gleitman, 1977; Seidenberg, 1985a). In an alphabetic writing system, such as the one used for English, the symbols—letters and letter clusters—generally correspond to phonemes. An advantage of this type of writing system is that it affords two ways to recognize words. First, the reader could recognize words on a visual basis, ignoring the fact that the symbols encode information about phonology. In this case, word recognition would be like other pattern recognition processes used in recognizing objects or nonalphabetic symbols. Having recognized the letter string as a token of a particular lexical type, the reader could then access its meaning, an outcome that has been termed *direct access* (Coltheart, 1978). Second, recognition could be based on a phonological code derived on the basis of the reader's knowledge of the correspondences between spelling and pronunciation. The meaning of a word could then be accessed using this derived phonological code, an outcome that has been termed *phonologically mediated access*. The extent to which readers of English recode letter strings into phonological representations to access meaning has been a central issue in research on word recognition (see Carr & Pollatsek, 1985; McCusker, Hillinger, & Bias, 1981, for reviews).

Early research on word recognition in English suggested that phonological information is always used to access meaning (Gough, 1972; Rubenstein, Lewis, & Rubenstein, 1971; Spoehr & Smith, 1973). An opposing view claimed that phonological information plays no role in skilled word recognition and emphasized the use of direct visual access (Becker, 1976, 1980; Brown, 1987; Johnson, 1975; Paap, Newsome, McDonald, & Schvaneveldt, 1982; Rumelhart & Siple, 1974; Smith, 1971). In light of subsequent research that provided considerable evidence for each of these positions, it is perhaps not surprising that many theorists proposed that both visual and phonological pathways exist and operate in parallel (Carr & Pollatsek, 1985; Coltheart, 1978, 1980; Forster & Chambers, 1973; Laberge & Samuels, 1974; Meyer, Schvaneveldt, & Ruddy, 1974; Paap, McDonald, Schvaneveldt, & Noel, 1987; Patterson & Morton, 1985; Seidenberg, 1985a; Seidenberg & McClelland, 1989; Shallice, Warrington, & McCarthy, 1983). The dual-route theories they proposed differ in terms of the types of knowledge representations and processes involved, the relative importance of the two routes in reading English, and other issues. However, the theories share the idea that processing occurs along both visual and phonological pathways in parallel and that each route determines meaning at least some of the time. This dual-route approach has dominated the field and has been widely accepted.

Recently, Van Orden and his colleagues (Van Orden, 1987; Van Orden, Johnston, & Hale, 1988; Van Orden, Pennington, & Stone, 1990) argued in favor of the earlier position that meaning is obtained from printed words using phonological representations. These researchers critically assessed the evidence thought to support the direct access process and provided new evidence supporting the use of phonology. The most recent version of their theory (Van Orden et al., 1990) includes both direct and phonological recognition pathways,

This article is based on a doctoral dissertation submitted to McGill University by Debra Jared and supervised by Mark S. Seidenberg. This research was supported by Grant A7924 from the Natural Science and Engineering Research Council of Canada (NSERC) to Mark S. Seidenberg. Debra Jared was supported by an NSERC Postgraduate Scholarship and by a scholarship from the Fonds pour la Formation de Chercheurs et l'Aide à la Recherche of Quebec.

Correspondence concerning this article should be addressed to Debra Jared, who is now at the Department of Psychology, McMaster University, 1280 Main Street West, Hamilton, Ontario, Canada L8S 4K1. Electronic mail may be sent to jared@mcmaster.ca.

as in traditional dual-route models; however, Van Orden et al. argued that the characteristics of the computation from orthography to phonology result in this route predominating over the direct route in the computation of semantic codes. Because many words are homophones (e.g., *seen–scene*), a verification process is then needed to associate one of the activated meanings with the input spelling pattern. Although the theory of Van Orden et al. (1990) is more specific about the nature of the representations and computations involved in word recognition, their view that meanings are primarily activated by phonological representations and that a spelling check follows is similar to the proposal of Rubenstein et al. (1971). Thus, if Van Orden is correct, then after 20 years of research on the question of whether the access of meaning is phonologically mediated, reading theory is essentially back to the same position.

One of the principal reasons why so much ambiguity remains concerning the role of phonology in word recognition is that most studies have relied on the tasks of naming words aloud or making lexical decisions (deciding whether a stimulus is a word or nonword). It has become clear over time that both tasks may be of limited use in addressing questions concerning the representations used to activate meaning. Naming, for example, might be accomplished by using knowledge of spelling–sound correspondences. Hence, subjects may be able to name words aloud without activating meaning, as in the case of nonwords. At one time, it was thought that this was possible only for words that obey the spelling–sound rules of the language; however, more recent models such as Seidenberg and McClelland's (1989), using a somewhat different way of representing spelling–sound knowledge, suggest that it is possible even for words that violate the rules (exception words, such as *have* and *give*). If words can be named on the basis of spelling–sound knowledge alone, the naming task will not necessarily provide evidence concerning the codes used to activate meaning.

It is also doubtful whether studies using the lexical-decision task necessarily provide evidence concerning the codes used to activate meaning. The extent to which phonological information influences lexical decisions varies as a function of the composition of the stimuli in an experiment (Davelaar, Coltheart, Besner, & Jonasson, 1978; Waters & Seidenberg, 1985). The result is that different sorts of lists give different answers as to whether meaning is activated by phonology. It is unclear which experimental conditions represent the ones that are relevant to conditions that prevail in normal reading.

The likely explanation of list-dependent effects is that lexical decision involves discriminating words from nonwords, and subjects establish different decision criteria depending on the difficulty of this discrimination within a particular set of stimuli (Balota & Chumbly, 1984; Besner, Davelaar, Alcott, & Parry, 1984; Gordon, 1983; Seidenberg & McClelland, 1989; Waters & Seidenberg, 1985). According to this view, responses can sometimes be based on the orthographic or phonological familiarity of the stimuli before meaning has been activated. This makes it difficult to use the task as a way of examining factors that affect the activation of meaning. One extreme view is that "the lexical decision task is a laboratory model of no natural process. It does not help us learn how people pronounce words or how we extract their meaning for it requires neither" (Baron, 1985, p. 706).

Van Orden (1987; Van Orden et al., 1988, 1990) conclusions concerning the role of phonology were based on the results of his studies using a semantic-decision task. Subjects were given a category, such as *an article of clothing*, and had to decide whether a target word, such as *suit*, is a member of the category. In contrast to naming and lexical decision, this task definitely requires the subject to consult the meaning of the target word. The goal, then, is to determine whether phonological factors affect performance on this task. The phonological factor manipulated in the Van Orden et al. studies was whether the target word was a homophone. Homophones are pairs such as *rose-rows*.

On critical trials, the subjects saw a category, such as *a flower*, and a target, such as *rows*, a homophonic foil for the correct category exemplar *rose*. If meaning is activated by phonology, a homophone target should activate meanings associated with both members of the pair. If the meaning of the unseen exemplar (e.g., *rose*) is available, it may be difficult for the subject to decide that the target (e.g., *rows*) is not a member of the category. In early studies using this methodology, Meyer and Ruddy (1973) and Meyer and Gutschera (1975) reported that *no* responses to the question "Is this word the name of a fruit?" were slower when the test word was a homophone of a fruit name (e.g., *pair*) than when it was an unrelated homophone (e.g., *tail*). Several other studies yielded similar results (e.g., Banks, Oka, & Shugarman, 1981; Ellison, 1975, cited in McCusker et al., 1981). These results seemed to implicate phonological activation of meaning.

However, Coltheart, Davelaar, Jonasson, and Besner (1977) pointed out a problem with the interpretation of these results. They argued that phonological processing may lag behind visual processing and may influence responding only on *no* decisions, which are typically longer than *yes* decisions. That is, phonological information may become available to influence decisions only after meaning has already been activated on a visual basis. Hence, evidence for phonological effects on *no* decision latencies cannot be taken as evidence that phonological activation normally contributes to the activation of meaning. Coltheart et al. argued that only phonological effects on *yes* decisions would provide unequivocal evidence that meaning is activated by phonological representations.

That is what Van Orden (1987) provided. He examined false positive, *yes* responses to homophone foils (e.g., a flower–*rows*) and unrelated controls matched to the homophone in terms of orthographic similarity (e.g., *robs*). He found that subjects made significantly more errors on homophone foils than on spelling controls and that this effect did not depend on the frequency of the homophone foil (e.g., *rows*), the stimulus the subject was shown. Van Orden argued that this was evidence against dual-route models in which phonological activation of meaning occurs for low-frequency words only (e.g., Seidenberg, 1985a). Furthermore, he found that the size of the effect depended on the frequency of the corresponding exemplar (e.g., *rose*) and on the orthographic similarity of the foil and the exemplar. Subjects were less likely to falsely accept the foil as a member of the category when the spelling of the actual exemplar was familiar (because it is a high-frequency word) and when it was not very similar to that of the foil. This provided evidence that subjects perform a spelling check on activated candidates. The orthographic similarity effect disappeared when words were presented briefly and then pattern-masked. The pattern mask was thought to prevent the spelling check.

In subsequent studies, Van Orden et al. (1988) attempted to determine whether the phonological representation responsible for these effects is computed before or after the activation of meaning. Coltheart (1980) argued that phonological codes are accessed postlexically, because he did not observe a difference between performance on homophone foils that are regular or irregular in terms of spelling–sound correspondences. Coltheart's logic was that only regular words can be named on the basis of nonlexical spelling–sound correspondences; naming irregular words requires lexical access. The observation that regular and irregular homophones produced similar effects suggested that phonological codes were accessed postlexically in both cases. However, this reasoning is questionable in light of recent models, such as that of Seidenberg and McClelland (1989), in which the correct phonological codes for both regular and irregular words can be produced by a single nonlexical mechanism.

Van Orden et al. (1988) chose another approach, comparing false-positive error rates for both word and pseudoword homophone foils and their respective controls. Pseudohomophones, such as *sute*, do not have meanings; hence, their phonological codes cannot be activated by semantic information. Both types of foils produced significantly more errors than spelling controls, and error rates to the two types of foils did not differ. In addition, *yes* response latencies that were false-positive errors (e.g., an article of clothing–*sute*) were not significantly different than *yes* response latencies for control exemplars (e.g., an article of clothing–*dress*) for either word or nonword foils. Van Orden et al. (1988) therefore concluded that phonology is computed rapidly enough to affect positive responses in a categorization task. They also suggested that their data did not provide evidence for an independent route from orthography to meaning.

The model Van Orden (1987) proposed to explain these findings is similar to that of Rubenstein et al. (1971). A visually presented letter string activates its phonological representation, which in turn activates a candidate set of meanings. A verification process (Becker, 1976, 1980; Paap et al., 1982; Rubenstein et al., 1971; Schvaneveldt & McDonald, 1981) is then conducted in which each meaning is associated with its spelling and these spellings are compared with the input stimulus. The spelling check is performed on entries in descending order of their activation levels until a match is found. False candidates (e.g., *rose*) will be activated if they share a phonological representation with the target (e.g., *rows*), causing higher error rates on homophone foils than on spelling controls. A false candidate is more likely to slip by the verification procedure if it is orthographically similar to the target or if its spelling is not familiar.

To summarize, the Van Orden (1987; Van Orden et al., 1988) research and other studies using the semantic-decision task provide evidence that phonological information does activate the meanings of words. This is important because the task is more directly relevant to the activation of meaning than others that have been used. However, the thrust of Van Orden's research is that the use of phonology is much more widespread than previously assumed; the title of the Van Orden et al. (1988) article asserts that "Word Identification in Reading Proceeds From Spelling to Sound to Meaning." Van Orden et al. (1990) criticized much of the evidence for direct activation of meaning from orthography and concluded that semantic codes are exclusively (Van Orden, 1987; Van Orden et al., 1988) or predominantly (Van Orden et al., 1990) activated by phonological representations. This stands in contrast to other views that assume a more prominent role for the direct visual route. In Coltheart's (1978) version of the dual-route model, processing along the visual route is almost always faster than along the phonological route, so that phonological mediation occurs only rarely in skilled reading. Seidenberg (1985a, 1985b) proposed a version of the dual-route model in which the contribution of phonological information to the activation of meaning depends on the time course of the recognition process. This view holds that phonological effects will be observed on relatively slowly recognized items, such as low-frequency words, but not on high-frequency words that are recognized more quickly. The studies of Van Orden et al. appear to provide evidence that neither of these dual-route positions is correct, insofar as phonology seems to play a greater role in the activation of meaning.

Claims by Van Orden and his colleagues (Van Orden, 1987; Van Orden et al., 1988, 1990) about the broad use of phonology to activate meaning are open to question. The basic issue concerns the range of conditions under which phonological activation of meaning is observed. One question is whether the phonological effects in the Van Orden et al. studies generalize beyond the conditions under which they were observed or whether they are related to specific aspects of the methodology. For example, the task involves giving subjects a category followed by an exemplar or foil. One potential problem is that the category names may prime potential exemplars. Many of the categories used in the Van Orden et al. studies have a small number of exemplars (e.g., *part of a horse's harness*). The subject might preactivate a small number of potential targets (e.g., *rein, bit*) and attempt to match the target to this verification list. If the candidates were themselves generated on a phonological basis (as in speech production), this might influence the false-positive rate for homophone targets. Van Orden et al. (1988) obtained production frequencies and measures of typicality for the category exemplars used in one experiment and found that these measures did not correlate with the rate of false-positive errors on homophone foils. A stronger test would be provided by using broader categories, such as *living thing* or *object*, which do not permit subjects to generate the target in advance (see Monsell, Doyle, & Haggard, 1989, who advocate this approach).

A second issue is whether phonological activation of meaning is related to word frequency, as suggested by McCusker et al. (1981), Seidenberg (1985a, 1985b), and others. Van Orden (1987, Experiment 3) evaluated this hypothesis in one experiment and found no effect of the frequency of the homophone foil on false-positive errors. This was interpreted as evidence against the view that high-frequency words are recognized directly on a visual basis. One reason why Van Orden may not have found a lower error rate for high-frequency homophone foils is that effects of foil frequency may have been canceled out by effects of exemplar frequency. There was a confound in his stimuli in that the low-frequency foils had primarily high-frequency exemplars (which were associated with fewer errors) and the high-frequency foils had primarily low-frequency exemplars (which were associated with more errors). To examine the role of homophone foil frequency, both it and exemplar frequency need to be manipulated factorially. A second reason for Van Orden's failure to find an effect of foil frequency may be that spelling controls were not used in this experiment; rather, controls were neither phonologically similar nor visually similar to the exemplar (e.g., the control for *rain* was *dust*). Thus, errors on homophone foils could have been due to either visual or phonological similarity to the exemplar, and it is possible that errors on high-frequency foils were due only to their visual similarity to the exemplar, whereas errors on low-frequency foils were due to both visual and phonological similarity to the foil. In order to attribute effects to phonological processes, then, it is imperative that performance on homophone foils be compared with that on spelling controls.

In summary, before we can accept Van Orden's conclusion that meaning is activated exclusively (Van Orden, 1987; Van Orden et al., 1988) or predominantly (Van Orden et al., 1990) by phonological representations, it is necessary to determine whether similar results obtain under a broad range of conditions (e.g., when the methodology is changed so as to eliminate potential priming from the category name). In addition, it is necessary to examine the effects of word frequency more closely. Because there is very little other evidence that skilled readers activate the meanings of high-frequency words using phonology, performance on these items needs to be examined carefully with reference to appropriate orthographic controls.

The experiments presented below examine responses to homophone and nonhomophone stimuli on semantic-decision tasks. In Experiment 1, the frequency of the homophone foil and the frequency of the exemplar are factorially manipulated, and performance on the foils is compared with that on spelling controls to determine whether evidence for phonological activation of the meanings of high-frequency words occurs, using the same procedure as Van Orden (1987). In Experiment 2, broad categories are used to examine whether the results obtained in Van Orden's work and in Experiment 1 are due to priming from the category name. Experiment 3 explores whether the use of phonological information is strategic by including a manipulation intended to discourage subjects from using it, and Experiment 4 examines whether a foil must not only sound like but also look like an exemplar to produce more false-positive errors. The results of these four experiments suggest that the effects Van Orden and colleagues reported are more limited than their studies suggest. Finally, two further studies provide additional evidence concerning phonological effects on correct *yes* responses to exemplars.

Experiment 1

The goal of Experiment 1 was to determine whether it is possible to replicate Van Orden's (1987; Van Orden et al., 1988) finding that subjects make more false-positive categorization errors when target words are homophones of a category exemplar than when they are spelled similarly to a category exemplar. In addition, the study examines whether these effects depend on the frequencies of the exemplar and foil. Van Orden (1987, Experiment 3) found an effect of exemplar frequency but not foil frequency; however, foil and exemplar frequency were confounded in the study, and control words were not matched to foils for orthographic similarity to the exemplar. Two levels of exemplar frequency (high and low) were crossed with three levels of foil frequency (high, low, and pseudohomophone), which produced six groups (see Table 1).

Two other category-target conditions were included in addition to the homophone foil (e.g., car part-*break*) and the spelling control (e.g., car part-*brave*) conditions mentioned earlier. The exemplar (e.g., car part-*brake*) was included in order to be able to compare false-positive response latencies on homophone foils with *yes* response latencies for the actual exemplars. Van Orden et al. (1988) argued that their finding of similar latencies in these two conditions indicates that the orthographic information concerning the exemplars did not contribute to the decision process. Had this information been available, it might have been expected to facilitate *yes* responses to exemplars. They compared false-positive response latencies for homophone foils with a mean *yes* latency for six other exemplars of the category in their first experiment and

Table 1
An Illustration of the Conditions Used in Experiment 1

Group		Category–target relation			
Foil frequency	Exemplar frequency	Exemplar	Foil	Spelling control	Homophone control
High	High	male relative-*son*	male relative-*sun*	male relative-*sin*	air vehicle-*sun*
High	Low	car part-*brake*	car part-*break*	car part-*brave*	painter's equipment-*break*
Low	High	child's toy-*ball*	child's toy-*bawl*	child's toy-*bail*	construction material-*bawl*
Low	Low	parasite-*flea*	parasite-*flee*	parasite-*flex*	award-*flee*
PW	High	footwear-*shoes*	footwear-*shews*	footwear-*shoss*	water vehicle-*shews*
PW	Low	canine-*fox*	canine-*focks*	canine-*fow*	fastner-*focks*

Note. PW = pseudoword.

with a single other exemplar in their second experiment. The actual exemplar is a better comparison because it controls for variables, such as typicality and production frequency, that affect the ease of making the category decision, and it is also a homophone.

In a fourth condition, subjects had to decide whether the homophone foil was a member of a completely unrelated category (e.g., painter's equipment-*break*). This allowed the comparison of rejection times and errors for homophones when they sounded as if they belonged to the category and when they did not. Banks et al. (1981) included a similar condition to control for the possibility that homophones could take longer to reject and be more prone to errors than spelling controls because they are represented differently from non-homophones and not because they sound like a member of the category. This homophone control condition was not included in Van Orden's (1987; Van Orden et al., 1988) experiments.

Thus, there are several comparisons of interest in the experiment. The most important of these are the comparisons of the false-positive error rate for homophone foils with those on spelling controls and homophone controls. If more false-positive errors are made on homophone foils than on either spelling controls or homophone controls, it would provide evidence that phonological information contributed to the activation of word meaning. Correct *no* response latencies for homophone foils were compared with those for spelling controls and for homophone controls. Van Orden (1987) and Van Orden et al. (1988) argued that similar *no* latency distributions provide evidence that all words undergo a spelling check, not just homophones. A problem with *no* latencies, however, is that—as Coltheart et al. (1977) suggested—they may include additional processing that is not done when a target is a member of the category; as such, the evidence from *no* latencies is much less important than evidence from false-positive errors. Finally, false-positive response latencies on homophone foils were compared with correct *yes* latencies on exemplars to examine the claim by Van Orden et al. (1988) that exemplars do not benefit from having orthographic information consistent with the *yes* response. The effect of the frequency of the exemplar and the frequency of the foil on the magnitude of these effects was also examined. Van Orden's (1987) view predicts that phonological effects will occur for both high- and low-frequency words, whereas dual-route models, such as Seidenberg's (1985a, 1985b), predict phonological effects for low-frequency words only. A comparison of the size of phonological effects for the word groups and the pseudoword groups provides additional evidence as to whether the effects arise before or after the activation of meaning.

The inclusion of the exemplar and homophone control conditions required a change in design from Van Orden's experiments. He presented all stimuli in a single session. To avoid intralist repetition effects, the present experiment was conducted in four sessions, each separated by at least 1 week. Only one member of a stimulus quadruple appeared in each session. For example, the trials car part-*brake* (exemplar), car part-*break* (homophone foil), car part-*brave* (spelling control), and painter's equipment-*break* (homophone control) all appeared in different sessions. As in Van Orden's experiments, a large number of filler trials were included, because several researchers have suggested that subjects change their processing strategies when a large proportion of the stimuli are homophones (Davelaar et al., 1978; Hawkins, Reicher, Rogers, & Peterson, 1976; McQuade, 1981).

Method

Subjects. Twelve McGill University undergraduates were paid $10 each to participate in the study. All were native speakers of English.

Stimuli. There were 288 experimental trials and 720 filler trials. The experimental trials consisted of 72 quadruples (see Appendix A). The first step in the construction of the quadruples was to choose 72 pairs of homophones. In order to be able to use the same stimuli with broad category names in Experiment 2, the homophone pairs were selected so that one member of each pair—the exemplar—was either a living thing or an object, and the other—the foil—was neither. Half of the exemplars were living things and half were objects. The 72 homophone pairs were chosen so that they fell into six groups of 12 pairs each. The six types of groups were produced by factorially manipulating the frequency of the exemplar (high and low) and the frequency of the foil (high, low, and pseudohomophone). The mean frequencies (from Kucera & Francis, 1967) of the exemplars and homophone foils for each of the six groups are presented in Table 2.

The second step was to choose a spelling control (e.g., car part-*brave*) for each of the 72 homophone foils. The spelling controls needed to be as similar to the exemplars in terms of orthography as the homophone foils were to the exemplars. To accomplish this, the orthographic similarity of each foil to its exemplar (e.g., *break, brake*) was calculated using Weber's (1970) graphic similarity measure (used by Van Orden, 1987; Van Orden et al., 1988), and then another stimulus was chosen that was as similar as possible to the exemplar (e.g., *brave, brake*) on this measure and was similar to the foil in frequency. The mean similarity scores of the foils to the exemplars and of the spelling controls to the exemplars for each of the six groups are presented in Table 2, along with the mean frequencies of the spelling controls.

In the final step, a category name was chosen for each of the 72 exemplars (e.g., car part-*brake*). These category names were also used for the matched homophone foils (e.g., car part-*break*) and spelling controls (e.g., car part-*brave*). The fourth member of each quadruple, the homophone control, consisted of the homophone foil and an unrelated category name (e.g., painter's equipment-*break*). To create these unrelated categories, the category names within each of the six groups were shuffled (e.g., car part-*cellar*, store personnel-*latter*, painter's equipment-*break*).

Four lists containing 72 experimental trials were created, with each member of a quadruple on a different list. Eighteen items from each of the category-target conditions (exemplar, homophone foil, spelling control, and homophone control) appeared on each list. These 18 consisted of 3 from each of the six exemplar frequency/foil frequency groups. The homophone controls were placed on lists such that no category name appeared more than once on a list. The number of experimental *yes* trials on each list was 18 (the exemplars), and the number of experimental *no* trials on each list was 54 (18 homophone foils, 18 spelling controls, and 18 homophone controls). The homophonic experimental stimuli were 21.5% of the stimuli on the entire list (16.7% were word homophones and 4.8% were pseudohomophones); one third of these were foil trials.

In addition to the experimental stimuli, 720 filler trials were included, 180 on each list. One hundred and eighty categories were

Table 2
Mean Word Frequency and Mean Orthographic Similarity to Exemplar for the Stimuli in the Six Experimental Conditions in Experiments 1–3

	Word frequency			Similarity to exemplar	
Group	Exemplar	Foil	Spelling control	Foil	Spelling control
HF foil/HF exemplar	85.3	83.4	76.5	.63	.64
HF foil/LF exemplar	7.3	86.4	88.4	.62	.63
LF foil/HF exemplar	92.1	6.9	5.9	.66	.70
LF foil/LF exemplar	5.0	4.5	4.5	.65	.66
PW foil/HF exemplar	83.8	—	—	.66	.67
PW foil/LF exemplar	6.8	—	—	.67	.68

Note. Word frequency was calculated using the Kucera and Francis (1967) norms, and orthographic similarity was calculated using Weber's (1970) measure. HF = high frequency; LF = low frequency; and PW = pseudoword. The statistics for the homophone control group are the same as for the foil group because the only difference in these conditions was in the preceding category name.

chosen, and each category appeared once on each list. Half of these were living-thing categories and half were object categories as required by Experiment 2. Four words were then chosen for each category, but not all of these were exemplars. Of the 180 filler words on each list, 108 were exemplars of their categories and 72 were not exemplars. This ensured that across each entire list (experimental trials plus fillers), there were the same number of *yes* and *no* trials (126). Because the homophone foils came from several different grammatical classes, the *no* filler words were also drawn from several grammatical classes so that this information would not provide a cue for the correct response. The *no* fillers were chosen so that the correct response would still be *no* when the categories were changed to *living thing* and *object* in Experiment 2.

In summary, there were four lists, each having an equal number of *yes* and *no* trials. The same 252 category names appeared on each list; the target appearing with a given category name was different on each list. The only targets that appeared twice were the 72 homophone foils that appeared with two different category names. The order of presentation within each list was pseudorandom, with no more than three *yes* or *no* trials appearing in succession.

An additional 16 categories and 16 stimuli were chosen for a practice list. Half of the stimuli were exemplars of their categories and half were not.

Procedure. The subjects completed four 30-min experimental sessions, which were separated by at least 1 week. In a session, the subjects first saw the 16 practice trials and then the 252 experimental trials from one of the lists. Subjects were given feedback on each practice trial to ensure that they understood the task. The order of presentation of the experimental lists was counterbalanced across subjects such that each list was presented three times in each of the four session positions. No subject saw any of the experimental lists twice.

On each trial a fixation point (*) appeared for 500 ms, followed by the category name, which was displayed for 2 s; the target stimulus then appeared and remained visible until the subject responded. Subjects were instructed to indicate whether the target stimulus was a member of the given category by depressing one of two telegraph keys as quickly as possible. The left key was used to indicate *no* responses, and the right key was used to indicate *yes* responses. The intertrial interval was 1.5 s.

Stimuli were presented in lowercase letters in the center of an IBM monitor (Model 5154) attached to an IBM PC-AT computer. A real-time clock in the computer calculated response times in milliseconds from the time the target stimulus appeared on the screen to the time the subject depressed one of the telegraph keys. The computer also recorded which key was depressed.

Results

Three sets of analyses were conducted. The first set examined error rates (false *yes* responses) on homophone foils, spelling controls, and homophone controls. A second set examined correct *no* latencies on these three groups. A third set compared erroneous *yes* latencies on homophone foils and correct *yes* responses on exemplars. Analyses of variance using both subject and item means (Clark, 1973) were used in the first 2 sets of analyses. There were three factors in each analysis: category–target relation (homophone foil, spelling control, and homophone control), target frequency (high, low, and pseudoword), and frequency of the exemplar (high and low). The factors were treated as within-subjects (repeated-measures) factors in the analyses using subject means and as between factors in the analyses using item means. Planned comparisons were performed to test for significant differences between pairs of means that were theoretically relevant. In this and subsequent experiments, subject means are reported in the text and figures.

False-positive error data. Percentage errors were arcsine-transformed before analysis, although the untransformed data produced essentially the same results. The untransformed percentages are reported in the text. There was a main effect of category–target relation, $F(2, 22) = 38.96$, $p < .001$, by subjects, and $F(2, 198) = 19.17$, $p < .001$, by items. Subjects made 16.2% errors on homophone foils (e.g., car part–*break*), 6.8% on spelling controls (e.g., car part–*brave*), and 3.1% on homophone controls (e.g., painter's equipment–*break*). Planned comparisons indicated that significantly more errors were made on homophone foils than on either spelling controls, $F(1, 11) = 48.10$, $p < .001$, by subjects, and $F(1, 198) = 16.75$, $p < .001$, by items, or homophone controls, $F(1, 11) = 45.39$, $p < .001$, by subjects, and $F(1, 198) = 36.82$, $p < .001$, by items.

The main effect of target frequency was significant in the subject analysis, $F(2, 22) = 9.94$, $p < .001$, and approached significance in the item analysis, $F(2, 198) = 2.93$, $p < .06$. Subjects made more errors on high-frequency words (11.5%) than on low-frequency words (8.5%) and pseudowords (6.2%). The interaction between category–target relation and target frequency was not significant by either subjects or items (both

Fs < 1). Consistent with Van Orden's (1987; Van Orden et al., 1988) findings, the differences between homophone foils and spelling controls were of similar magnitude for high-frequency words (8.3%), low-frequency words (10.6%), and pseudowords (9.0%). The differences between homophone foils and homophone controls were also of similar magnitude for high-frequency words (14.6%), low-frequency words (13.6%), and pseudowords (11.1%).

The main effect of exemplar frequency was not significant, $F(1, 11) = 1.00$, $p > .05$, by subjects ($F < 1$, by items). The interaction of category–target relation and exemplar frequency was significant by subjects, $F(2, 22) = 9.05$, $p < .01$, but not by items, $F(2, 198) = 2.24$, $p > .05$. The difference in percentage errors between homophone foils and spelling controls was 14.1% for those with low-frequency exemplars and 4.6% for those with high-frequency exemplars. Consistent with Van Orden's (1987) findings, the planned comparisons indicated that the difference for foils with low-frequency exemplars was significant, $F(1, 11) = 101.50$, $p < .001$, by subjects, and $F(1, 198) = 17.96$, $p < .001$, by items, but the difference for foils with high-frequency exemplars was not significant. The difference between homophone foils and homophone controls was 17.4% for those with low-frequency exemplars and 8.8% for those with high-frequency exemplars.

The triple interaction was significant by subjects, $F(4, 44) = 3.09$, $p < .05$, but not by items, $F(4, 198) = 1.00$, $p > .05$. Planned comparisons performed to examine the difference in error rate between homophone foils (e.g., car part-*break*) and spelling controls (e.g., car part-*brave*) in each of the six groups revealed that the difference was significant by both subjects and items only for low-frequency foils with low-frequency exemplars, $F(1, 11) = 26.25$, $p < .001$, by subjects, and $F(1, 198) = 8.44$, $p < .01$, by items, and pseudohomophones with low-frequency exemplars, $F(1, 11) = 34.33$, $p < .001$, by subjects, and $F(1, 198) = 7.69$, $p < .01$, by items. These data are summarized in Figure 1.

Planned comparisons were also performed on the differences between homophone foils (e.g., car part-*break*) and homophone controls (e.g., painter's equipment-*break*). These differences were significant by subjects and items in four of the six groups: high-frequency foils with high-frequency exemplars (16.7%), high-frequency foils with low-frequency exemplars (12.5%), low-frequency foils with low-frequency exemplars (22.9%), and pseudohomophone foils with low-frequency exemplars (16.7%), all ps < .02. The differences for low-frequency foils (4.2%) and pseudohomophone foils (5.5%) with high-frequency exemplars were not significant.

No *reaction time data.* This analysis examined correct *no* latencies on homophone foils (e.g., car part-*break*), spelling controls (e.g., car part-*brave*), and homophone controls (e.g., painter's equipment-*break*). A subject's response latency on a trial was only included in the analyses if the subject responded correctly on that trial and also responded correctly to the other three members of the stimulus quadruple (i.e., responses on car part-*brake*, car part-*break*, car part-*brave*, and painter's equipment-*break* all had to be correct for them to be included). This procedure ensured that the same number of scores were included in each of the three category–target relation conditions in the analysis; the same procedure was used by Van Orden (1987) and Van Orden et al. (1988). The additional constraint that a trial was included only if the subjects correctly responded to the related exemplar ensured that a *no* response to the foil (e.g., car part-*break*) occurred because subjects correctly avoided confusion with the exemplar and not because subjects activated the representation of the exemplar (e.g., *brake*) using phonology, but thought that the exemplar was a nonmember of the category (e.g., that a *brake* was not a car part). These criteria were met by 69.6% of responses. Another 21.6% of responses were correct but were discarded because an error was made on another member of the quadruple. Essentially the same results were obtained when all correct responses were included in the anal-

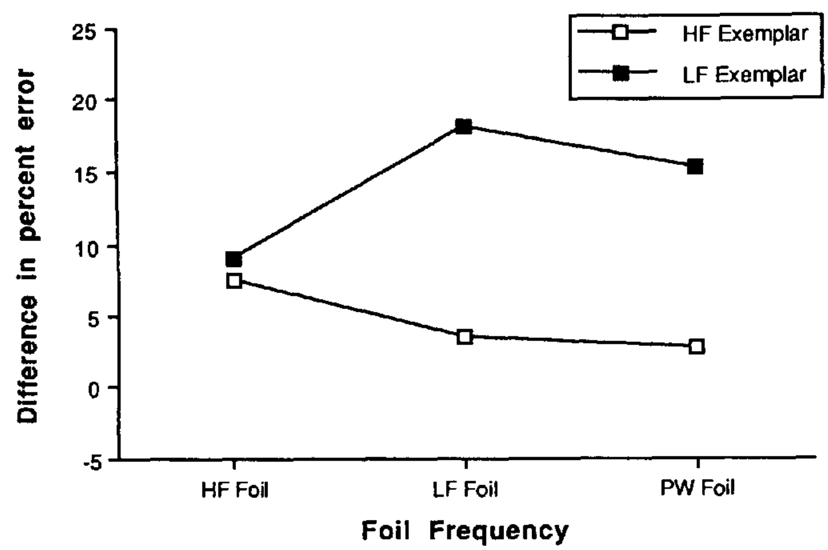

Figure 1. The difference in mean percentage errors between homophone foils and spelling controls in Experiment 1. (HF = high frequency; LF = low frequency; and PW = pseudoword.)

yses. Nine response times greater than 1,500 ms were replaced with times of 1,500 ms.

There was a main effect of category–target relation in the *no* latency data, $F(2, 22) = 18.61$, $p < .001$, by subjects, and $F(2, 195) = 10.53$, $p < .001$, by items. Planned comparisons revealed that homophone foils (702 ms) produced significantly longer latencies than spelling controls (665 ms), $F(1, 11) = 15.92$, $p < .01$, by subjects, and $F(1, 195) = 7.61$, $p < .01$, by items, and significantly longer latencies than homophone controls (641 ms), $F(1, 11) = 37.72$, $p < .001$, by subjects, and $F(1, 195) = 20.77$, $p < .001$, by items. None of the other main effects or interactions were significant.

Van Orden (1987) and Van Orden et al. (1988) also observed a main effect of category–target relation in their *no* latency data, but argued that it was due to outlying scores. To examine whether this was the case in the present data, two further analyses were performed using cutoff values of 1,000 ms and 850 ms. Scores greater than the cutoff value and the corresponding scores in the other two category–target relation conditions were removed from the analyses. This resulted in the removal of 10.5% of the scores included in the original analysis when the cutoff was 1,000 ms and 32.4% when the cutoff was 850 ms. With a 1,000-ms cutoff value, the main effect of category–target relation was still significant, $F(2, 22) = 21.14$, $p < .001$, by subjects, and $F(2, 195) = 13.59$, $p < .001$, by items. Planned comparisons revealed that homophone foils produced significantly longer latencies than either spelling controls (19 ms), $F(1, 11) = 15.82$, $p < .01$, by subjects, and $F(1, 195) = 7.40$, $p < .01$, by items, or homophone controls (46 ms), $F(1, 11) = 31.18$, $p < .001$, by subjects, and $F(1, 195) = 27.19$, $p < .001$, by items. With an 850-ms cutoff value, the main effect of category–target relation approached significance by subjects, $F(2, 22) = 3.29$, $p < .06$, and was significant by items, $F(2, 192) = 5.19$, $p < .01$. However, there was no longer a difference between homophone foils and spelling controls (0 ms). The difference between homophone foils and homophone controls (18 ms) approached significance by subjects, $F(1, 11) = 4.37$, $p = .06$, and was significant by items, $F(1, 192) = 9.00$, $p < .005$.

Yes *response latencies.* In a final set of analyses, correct *yes* response latencies on exemplars were compared with false-positive response latencies on homophone foils to examine whether exemplars benefit from having orthographic information consistent with the *yes* response. One-tailed t tests of subject and item means were used because too few errors were made to perform an analysis of variance with the foil frequency and exemplar frequency factors. An item was included in these analyses only if a subject made both a false-positive error on the homophone foil and correctly accepted the matched exemplar. Nine out of 140 false-positive latencies on homophone foils were not included because errors were made on the corresponding exemplars. Subjects responded *yes* significantly faster to exemplars (546 ms) than to the homophone foils (620 ms), $t(11) = 3.61$, $p < .005$, by subjects, and $t(88) = 2.33$, $p < .02$, by items. This tendency occurred in each of the six groups; however, there were too few errors in each group (range: 8–33) to examine the effects of exemplar and foil frequency. The difference between exemplars and homophone foils was exaggerated by a few extreme scores. When 12 (9.2%) false-positive homophone foil responses longer than 900 ms and the corresponding exemplar responses were removed, the difference between correct *yes* responses (540 ms) and false-positive response latencies (566 ms) only approached significance by subjects, $t(11) = 1.78$, $p < .06$, and was not significant by items, $t(82) = 1.00$, $p > .05$.

Discussion

The results of the error analyses were similar to those of Van Orden (1987) and Van Orden et al. (1988). They found that subjects produced significantly more false-positive errors on homophone foils than on spelling controls, a finding replicated here. Furthermore, Van Orden (1987, Experiment 3) did not observe an effect of foil frequency on the magnitude of the difference between homophone foils and controls when all of the stimuli were words, and Van Orden et al. (1988) observed a similar-sized difference between foils and spelling controls for words and pseudowords. Consistent with these findings, in Experiment 1, the size of the difference between homophone foils and spelling controls was similar for high-frequency words, low-frequency words, and pseudohomophones. Van Orden (1987) found that the magnitude of the difference in false-positive error rates between homophone foils and spelling controls was influenced by the frequency of the exemplar. Consistent with this, in Experiment 1, the difference between homophone foils and spelling controls was larger when homophone foils had low-frequency exemplars than when they had high-frequency exemplars.

In the *no* latency data, an overall difference between homophone foils and spelling controls was observed, as in the Van Orden (1987) and Van Orden et al. (1988) studies. Van Orden et al. argued that these effects were due to outliers in the homophone foil group and that otherwise the *no* latency distributions for the two groups were essentially the same. Van Orden (1987) did not report the percentage of scores that were removed from the analysis that produced similar means for the two groups, but Van Orden et al. (1988) reported that the means did not differ when about 30% of scores were removed. Here, the difference between homophone foils and spelling controls remained when 10.5% of scores were removed, a number that could reasonably be considered for outliers. As in Van Orden et al. (1988), the difference disappeared when 32.4% were removed. However, when this many scores are removed, the analysis is based on only 47.4% of critical *no* trials (because only 69.6% were included in the original analysis). This figure was approximately 40% in the experiment by Van Orden et al. (1988). It is debatable whether meaningful conclusions can be based on the data from such a small percentage of critical trials.

Van Orden et al. (1988, Experiment 1) found similar latencies for false-positive responses on homophone foils and for correct *yes* responses on category exemplars. In the second experiment in the article, the exemplars were matched more closely to the foils, and this time the *yes* latencies were faster for exemplars than for foils (49 ms faster for words and 63 ms faster for pseudowords), although the differences were not significant. Van Orden et al. argued that the differences were due to outlier latencies for homophone foils and demonstrated that the means for foils and exemplars were the same when 30% of scores were removed. In Experiment 1, *yes* latencies collapsed across groups were significantly faster for exemplars than for homophone foils, but there was no difference when about 10% of trials with the longest latencies were removed. This suggests that the exemplars benefited little from having orthographic information consistent with the *yes* response. However, a problem with this conclusion is that it is based on a small number of responses. Another problem is that it may not be valid to compare correct and erroneous *yes* responses because erroneous *yes* responses may be made before foils are fully processed, whereas correct responses may be made when the processing of the exemplar is complete.

In summary, the overall results of Experiment 1 replicated Van Orden's results quite well. In addition, Experiment 1 indicates that the higher error rate for homophone foils than for spelling controls is not due to a more general difficulty in processing homophones, because homophone foils produced significantly more errors than homophone controls. The difficulty with homophones on a semantic-decision task arises only when they sound like an exemplar of a category.

Van Orden argued that the results of his experiments provided evidence that the meanings of words are exclusively (Van Orden, 1987; Van Orden et al., 1988) or predominantly (Van Orden et al., 1990) activated by their phonological representations. The findings cited in support of this conclusion were the higher error rate on homophone foils than on spelling controls, regardless of frequency; the similar *yes* latencies on exemplars and foils; and the similar false-positive error rates on homophone and pseudohomophone foils. This last finding provides evidence that the phonological representation responsible for these effects is computed before, rather than after, the activation of meaning. Similar *no* latency distributions for homophone foils and spelling controls suggested to Van Orden that subjects were making use of a spelling check for both types of stimuli. He interpreted the effect of exemplar frequency on homophone foil error rates as additional support for the verification hypothesis, because better knowledge of the exemplar would facilitate a spelling check. The time-course dual-route theory, he claimed, predicts that a difference between homophone foils and spelling controls should be observed for low-frequency foils but not for high-frequency foils. A high-frequency foil activates its meaning directly and thus should not be influenced by its phonological representation, but the meaning of a low-frequency word is activated by the phonological route and so should be more susceptible to phonological confusion errors. However, in neither Van Orden's (1987) experiment nor in Experiment 1 was the difference between homophone foils and spelling controls affected by foil frequency. In addition, Van Orden pointed out that dual-route theory predicts that the difference between homophone foils and spelling controls should be greater when homophone foils have high-frequency exemplars than when they have low-frequency exemplars. This is because the high-frequency exemplars would reach maximum levels of activation sooner than low-frequency words (Morton, 1969) and thus would be more likely to be mistakenly selected. In contrast, in Van Orden's experiment and in Experiment 1, fewer errors were made on homophone foils with high-frequency exemplars than on those with low-frequency exemplars.

The overall results of Experiment 1 appear to support Van Orden's (1987) view. Closer inspection of the data, however, indicates that the evidence for phonological activation of the meanings of high-frequency words was not strong. Although high-frequency foils produced significantly more errors than spelling controls, when the high-frequency foils with high-frequency exemplars and high-frequency foils with low-frequency exemplars were examined separately, the differences for each group were only significant in the subjects analysis, which suggests the effect is limited to only some of the words. In addition, the magnitude of the difference between homophone foils and spelling controls was similar for high-frequency foils with high-frequency exemplars (7.6%) and high-frequency foils with low-frequency exemplars (9.0%). This is not consistent with Van Orden's proposal that all words undergo a spelling check, because in his view, subjects should have more complete knowledge of the spelling of a high-frequency exemplar and thus should be better able to detect its homophone foil in the spelling check.

The time-course dual-route theory (Seidenberg, 1985a, 1985b), on the other hand, has difficulty explaining the absence of a false-positive error effect for low-frequency and pseudohomophone foils with high-frequency exemplars. These stimuli should have activated semantic representations on the basis of phonological information, and because the semantic representation of the exemplar is also activated with homophone foils, the foils should have been more susceptible to false-positive errors than spelling controls.

The weak evidence for phonological activation of the meanings of high-frequency words may be due to the small number of stimuli used in each of the groups in the experiment (although more were used here than in Van Orden's experiments), and perhaps if more high-frequency homophone stimuli were available, stronger effects would be found. However, another possibility is that the effect of homophone foils versus spelling controls was exaggerated in Experiment 1 and in Van Orden's experiments, particularly for high-frequency words, because of priming from the category name. Experiment 2 explored this possibility.

Experiment 2

Experiment 2 addressed whether the false-positive error rate in the category-decision task is affected by task-specific strategies. Specifically, false-positive errors may arise, in part, from subjects' attempts to generate potential targets. Balota and Chumbly (1984), Forster (1988), and Monsell et al. (1989) suggested that category names may prime exemplars in a category-decision task. Becker's (1976) verification model specifically holds that subjects generate a semantic-candidate set when shown a prime. Several studies have demonstrated priming by category names in a lexical-decision task (for a review, see Neely, 1991). Using a variation of the semantic decision task, Rosch (1975) found that prior presentation of the category name facilitated judgments of whether a pair of words belonged to the same category relative to neutral prime (the word *blank*) and that the size of the facilitation effect was similar for pairs of words that were good and poor exemplars of the category. Further evidence for the priming of exemplars by category names in a semantic-decision task comes from the pilot work of Van Orden (1987, Experiment 2), who observed that subjects needed to view an exemplar target for less time than a nonexemplar target in order to be able to report it.

Once subjects are shown a category name, they may begin to generate possible semantic candidates, and these candidates may themselves activate corresponding phonological representations, much as when a spoken response is prepared. Consequently, when the target stimulus appears on the screen, there may already be considerable activation in both the semantic and phonological systems. The phonological representation of the target word could become available on at least two bases that do not involve phonological activation of meaning. First, the phonological representation could be activated after the semantic representation. Second, the phonological representation could be activated on the basis of an independent computation from orthography to phonology. The combination of activation of the phonological representation by the prime and by the target may be enough to trigger a *yes* response, which would be a false-positive response in the case of a homophone foil. In neither case would the false-positive errors on homophone foils reflect the use of phonological information to activate meaning.

Van Orden's (1987; Van Orden et al., 1988) strong view that meaning is activated by phonological representations predicts that the effects that he observed should not be dependent on the nature of the category names used. With broad categories, subjects will still have to activate the meanings of the target words to perform the task; therefore, there should still be more errors on homophone foils than on spelling controls for both high- and low-frequency targets. His position would be called into question if the effects were highly dependent on this task-specific aspect of the studies.

To reduce the likelihood that phonological representations are activated by the category name before the presentation of the target, two broad categories, *living thing* and *object*, were used in Experiment 2, instead of the more specific categories used in previous studies. This variant of the task preserves the important feature that subjects must activate the meanings of the targets in order to make their decisions but reduces the probability of predicting the target in advance. Exactly the same target stimuli as in Experiment 1 were used to ensure that any differences between the experiments could be attributed to effects of category specificity. If the false-positive errors found in Experiment 1 reflect only the activation of meaning by phonological representations, then results of Experiment 2 should be similar. If the results were due to priming from the category name, fewer, if any, false-positive errors should be observed in Experiment 2.

Method

Subjects. Twelve McGill University undergraduates were paid $10 each to participate in the study. All were native speakers of English. None had participated in Experiment 1.

Stimuli. The target stimuli were the same 1,008 as those used in Experiment 1 (see Appendix A). The category names used were *living thing* and *object*. Half of the experimental stimuli and half of the filler stimuli were preceded with the category *living thing*, and half of each were preceded by the category *object*. The category name that preceded a target was chosen so that the correct response to a target stimulus was the same as in Experiment 1. Half of the items in each category had a correct response of *yes*, and half had a correct response of *no*. The target items appeared on the same one of four lists and in the same order as in Experiment 1. No more than three trials with the same category name or the same correct response appeared in succession.

Procedure. The viewing conditions and procedure were exactly the same as in Experiment 1.

Results

The data were analyzed in the same manner as in Experiment 1. As before, three sets of analyses were conducted. The first set examined error rates (false *yes* responses) on homophone foils, spelling controls, and homophone controls; the second set examined correct *no* latencies on these three groups; and the third set compared erroneous *yes* latencies on homophone foils and correct *yes* responses on exemplars. The scores on six words in the homophone control condition (*sun*, *mail*, *beach*, *cellar*, *prints*, and *buoy*) were not included in the analyses, for the following reason: The homophone control condition consisted of pairing the homophone foil words with an unrelated category name, such that the correct response was also *no* (e.g., in Experiment 1, a foil trial was car part–*break* and the homophone control was painter's equipment–*break*). However, although these six words are foils for the category *living thing*, they are also exemplars of the only other category used in the experiment (*object*) and thus could not be presented in an unrelated category. These words were included in the experiment anyway due to the difficulty of finding enough pairs of homophones in which one was a living thing or an object and the other one was neither. Because the main contrast of interest is between homophone foils and spelling controls, the exclusion of these items has little effect on the conclusions.

False-positive error data. The overall error rate (9.8%) was similar to that in Experiment 1 (8.7%). There was a main effect of category–target relation, $F(2, 22) = 12.57$, $p < .001$, by subjects, and $F(2, 192) = 6.30$, $p < .01$, by items. Subjects made 14.2% errors on homophone foils (e.g., object–*break*), 8.8% on spelling controls (e.g., object–*brave*), and 6.5% on homophone controls (e.g., living thing–*break*). Planned comparisons indicated that significantly more errors were made on homophone foils than on either spelling controls, $F(1, 11) = 15.47$, $p < .01$, by subjects, and $F(1, 192) = 6.96$, $p < .01$, by items, or homophone controls, $F(1, 11) = 13.67$, $p < .01$, by subjects, and $F(1, 192) = 11.39$, $p < .001$, by items.

The main effect of target frequency was not significant, $F(2, 22) = 1.53$, $p > .05$, by subjects ($F < 1$, by items). Subjects made 11.2% errors on high-frequency words, 9.2% on low-frequency words, and 9.1% on pseudowords. The interaction of category–target relation and target frequency was also not significant, $F(4, 44) = 2.39$, $p > .05$, by subjects

($F < 1$, by items). The difference between homophone foils and spelling controls was only 0.4% for high-frequency words, contrary to the significant effect in Van Orden (1987) and Experiment 1. However, for low-frequency words and pseudowords, the magnitude of the difference—8.0% in both cases—was similar to that found in Experiment 1. The difference between homophone foils and homophone controls was 4.6% for high-frequency words, 10.3% for low-frequency words, and 16.6% for pseudowords.

The main effect of exemplar frequency was not significant, $F(1, 11) = 3.06, p > .05$, by subjects ($F < 1$, by items). The interaction of category–target relation and exemplar frequency was marginally significant by subjects, $F(2, 22) = 3.32, p = .05$, but not by items ($F < 1$). Consistent with the results of Van Orden (1987) and Experiment 1, planned comparisons indicated that there was a significant difference between homophone foils with low-frequency exemplars and spelling controls (7.3%), $F(1, 11) = 14.42, p < .01$, by subjects, and $F(1, 192) = 4.69, p < .05$, by items, but the difference between homophone foils with high-frequency exemplars and spelling controls (3.5%) was not significant. The difference between homophone foils and homophone controls was 10.6% for those with low-frequency exemplars and 4.8% for those with high-frequency exemplars.

The triple interaction was significant by subjects, $F(4, 44) = 2.77, p < .05$, but not by items ($F < 1$). Planned comparisons were performed to examine the difference in error rate between homophone foils (e.g., object–*break*) and spelling controls (e.g., object–*brave*) in each of the six exemplar frequency/foil frequency groups. These differences are presented in Figure 2. As in Experiment 1, the tests revealed that the differences between homophone foils and spelling controls were significant for low-frequency words with low-frequency exemplars, $F(1, 11) = 15.75, p < .01$, by subjects, and $F(1, 192) = 4.93, p < .05$, by items, and for pseudowords with low-frequency exemplars, $F(1, 11) = 15.26, p < .01$, by subjects, and $F(1, 192) = 3.37, p < .07$, by items. None of the differences for the high-frequency word conditions approached significance (all $ps > .20$).

Also of interest are comparisons between homophone foils (e.g., object–*break*) and homophone controls (e.g., living thing–*break*). Planned comparisons indicated that this difference was significant for the same two groups, the low-frequency words with low-frequency exemplars (16.6%), $F(1, 11) = 38.74, p < .001$, by subjects, and $F(1, 192) = 7.15, p < .01$, by items, and the pseudowords with low-frequency exemplars (12.4%), $F(1, 11) = 15.26, p < .01$, by subjects, and $F(1, 192) = 5.12, p < .05$, by items. The differences for high-frequency foils with low-frequency exemplars (2.8%) and for high-frequency foils (6.3%), low-frequency foils (4.0%), and pseudohomophone foils (4.2%) with high-frequency exemplars were not significant.

No *reaction time data.* The correct *no* decision latencies on homophone foils, spelling controls, and homophone controls were included in this analysis. As in Experiment 1, a subject's response latency for an item was only included if the subject responded correctly on that trial and responded correctly to the other three members of the stimulus quadruple (e.g., object–*brake*, object–*break*, object–*brave*, and living thing–*break*). This criterion was met by 67.4% of responses. Another 22.2% of responses were correct but were discarded because an error was made on another member of the quadruple. Essentially the same results were found, however, when all correct reaction times were included. Twenty-five response times greater than 1,500 ms were replaced with times of 1,500 ms.

The main effect of category–target relation in the *no* latency data was significant by subjects, $F(2, 22) = 3.91, p < .05$, but not by items, $F(2, 192) = 1.46, p > .05$. The difference between homophone foils and spelling controls was 11 ms, and the difference between homophone foils and homophone controls was 34 ms. The main effect of target frequency approached

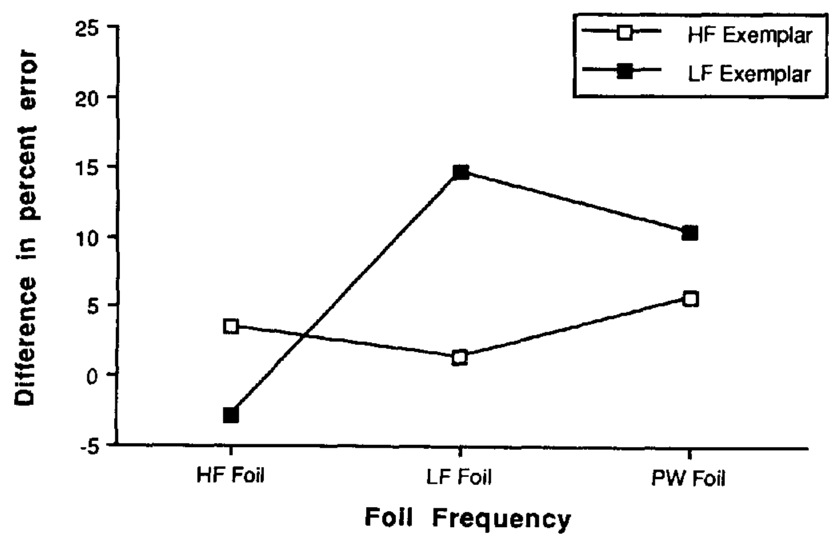

Figure 2. The difference in mean percentage errors between homophone foils and spelling controls in Experiment 2. (HF = high frequency; LF = low frequency; and PW = pseudoword.)

significance by subjects, $F(2, 22) = 3.26$, $p < .06$, and was significant by items, $F(2, 192) = 4.33$, $p < .02$. Subjects responded more quickly to high-frequency words (763 ms) than to low-frequency words (786 ms) and pseudowords (798 ms). None of the other interactions approached significance (all Fs < 1).

Yes *response latencies.* An item was included in these analyses only if a subject made both a false-positive error on the homophone foil and correctly accepted the matched exemplar. Ten out of 123 false-positive latencies on homophone foils were not included because the subjects failed to respond correctly to the corresponding exemplar. Subjects responded *yes* significantly faster to exemplars (678 ms) than to the homophone foils (775 ms) in the analysis by subjects, $t(10) = 2.94$, $p < .02$, but the difference was not significant by items, $t(80) = .99$, $p > .05$. This trend occurred in each of the six groups. As in Experiment 1, there were too few errors in each group (range: 11–29) to include frequency of exemplar and frequency of foil in the analysis.

Discussion

A significant overall difference in the false-positive error rate between homophone foils and spelling controls was found in Experiment 2, replicating the findings of Experiment 1 and those of Van Orden (1987; Van Orden et al., 1988). In addition, homophone foils produced significantly more errors than homophone controls, which—as in Experiment 1—indicates that the elevated rate of errors on homophone foils arises because they sound like a member of the category, not because homophones are generally harder to process.

However, unlike Experiment 1 and Van Orden's (1987) Experiment 3, high-frequency homophone foils did not produce more false-positive errors than spelling controls. Thus, the presence of an effect for high-frequency words depends on the type of category given to the subjects (see Figure 3). When the category was specific (Experiment 1), such as *car part*, subjects made significantly more errors (8.3%) on homophone foils than on spelling controls. However, when the category was broad (Experiment 2), such as *object*, the difference was only 0.4% on exactly the same words. This suggests that in Experiment 1, the effects for high-frequency words were inflated by priming from the category name. Subjects may have responded on the basis of a match between a phonological candidate activated by the category name and phonological activation generated by the target independent of meaning. The results of Experiment 2 suggest, then, that the use of specific categories in a category-decision task should be avoided when studying the activation of meaning in single-word reading so that subjects cannot make strong predictions about subsequent targets.

The failure to find a difference in false-positive error rates between homophone foils and spelling controls for high-frequency words suggests that meaning is not activated by phonology for these words. If the meanings of high-frequency words had been phonologically activated, two meanings would have been available for the homophones, whereas only one would have been available for the spelling controls. On at least some trials, subjects could be expected to choose the wrong homophone alternative and make a false-positive error. However, there was no evidence that the decision was more difficult for homophones than for spelling controls.

Although Experiment 2 provided no evidence that phonological information contributes to the activation of the meanings of high-frequency words, the results do suggest that it contributes to the activation of the meanings of low-frequency words. More false-positive errors were made on homophone foils than on spelling controls for low-frequency words and for pseudowords, which suggests that two meanings were available for these stimuli. The size of the difference was 8.0%

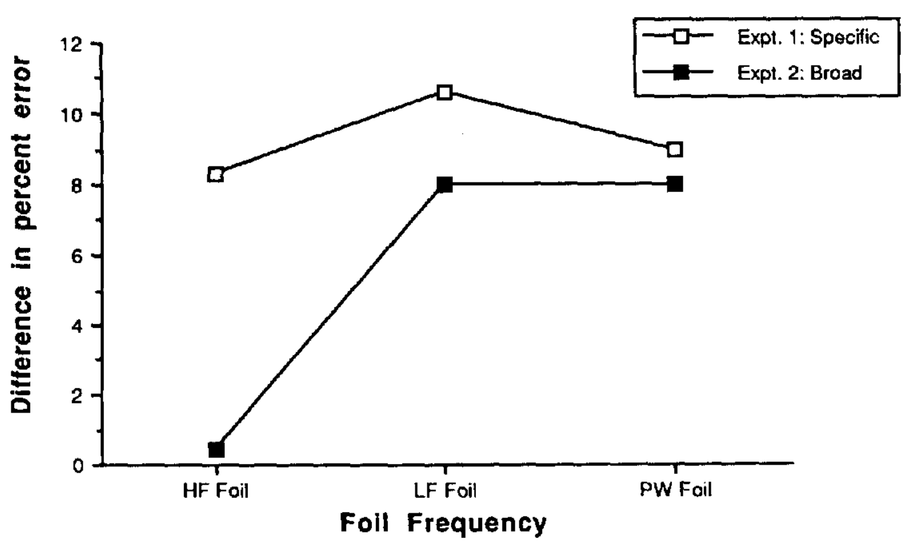

Figure 3. The effect of category specificity on the difference in mean percentage errors between homophone foils and spelling controls (collapsed across exemplar frequency). (HF = high frequency; LF = low frequency; PW = pseudoword.)

for both groups. In Experiment 1, the difference was 10.6% for low-frequency words and 9.0% for pseudowords. Van Orden et al. (1988) argued that similar-sized effects for words and pseudowords indicated that the phonological representation responsible for these effects is computed before, rather than after, the activation of meaning.

Two other findings were consistent with Van Orden's account of a spelling-check procedure for phonologically activated candidates. The size of the difference in the error data was larger for foils with low-frequency exemplars than for those with high-frequency exemplars. According to Van Orden's view, the lower error rate on foils with high-frequency exemplars occurs because subjects have better knowledge of the exemplar spellings and thus are better able to detect homophone foils. The second finding was that there was no difference between homophone foils and spelling controls in *no* latencies. Van Orden (1987) interpreted the lack of difference in the *no* latency data as indicating that all phonologically activated candidates, not just homophones, undergo a spelling check.

Van Orden et al. (1988) also claimed that their results indicate that orthographic information does not contribute to the activation of meaning. In their experiments, exemplars did not benefit from having orthography consistent with a *yes* response. In Experiment 2, the difference between *yes* latencies for exemplars and homophone foils was significant by subjects. However, because the *yes* analyses are based on a small number of responses, it would not be safe to conclude from these data that orthography never contributes to semantic activation.

Another piece of evidence that Van Orden (1987) used to support his view that candidates are activated exclusively by their phonological representations comes from his Experiment 2, which used tachistoscopic presentation. In that experiment, he found a large difference in errors between foils and spelling controls that was not dependent on the orthographic similarity of the foil to its exemplar. Van Orden argued that the mask that followed the brief presentation of the stimuli prevented the spelling check from being performed on phonologically activated candidates. The finding that priming from specific category names exaggerated errors on homophone foils in Experiment 1 suggests that when the spelling check was prevented by masked tachistoscopic presentation, the rate of false-positive errors on foils should have been considerably larger than 50% if candidates are activated exclusively by their phonological representations (Coltheart, 1978). That is, given two candidates activated by a phonological representation (the foil and the exemplar) and no orthographic information on which to tell them apart, subjects should be more likely to choose the exemplar because it is primed by the category name. Van Orden (1987, Experiment 2) found an error rate of only 43% on homophone foils, which indicates that there is some activation of candidates by orthographic information.

In summary, Experiment 2 indicates that evidence for phonological activation of the meanings of high-frequency words is not obtained when broad categories are used to reduce priming from the category name. There was evidence for phonological activation of the meanings of low-frequency words and evidence that a spelling check is performed on activated candidates.

Experiment 3

The results of Experiment 2 suggest that skilled readers make use of phonological information to activate the meanings of low-frequency words but not high-frequency words, as previously suggested by McCusker et al. (1981); Andrews (1982); Seidenberg, Waters, Barnes, and Tanenhaus (1984); Seidenberg (1985a, 1985b); and others. Experiment 3 explores the possibility that phonological activation of the meanings of low-frequency words is not obligatory, but rather is a strategy under the control of the subject.

Several authors (Coltheart, 1978; Davelaar et al., 1978; Hawkins et al., 1976; McQuade, 1981) proposed that the use of phonological information in visual word recognition is strategic and that subjects will avoid the use of phonology when it impairs performance on a task. They attempted to observe strategic use of phonology by varying the proportion of homophonic stimuli in lexical-decision experiments. The logic of the approach is that if there is phonological activation of meaning, subjects should make more errors or take longer to make decisions on homophonic (e.g., *bare*, *grone*) than nonhomophonic stimuli (e.g., *bake*, *grobe*). However, when many homophonic stimuli are included, subjects would notice that the phonological strategy was causing them to make a large number of errors, and so they would abandon it in favor of a visual strategy, in which case no effect of homophony would be expected. Using a lexical-decision task, Davelaar et al. (1978) found an effect of homophony for low-frequency words when the pseudoword distractors did not sound like English words (e.g., *slint*), but the homophone effect disappeared when the distractors were pseudohomophones (e.g., *grone*). Also using lexical decision, McQuade (1981) found a larger pseudohomophone effect (longer latencies for pseudohomophones compared with nonpseudohomophones) when a low proportion of the stimuli were pseudohomophones. Hawkins et al. (1976) found an effect of homophony in a tachistoscopic word-recognition task only when the stimulus list contained a low proportion of homophones. These researchers all concluded that subjects make use of a phonological strategy when this strategy causes few errors (e.g., when there are few homophonic stimuli), but abandon it and use a visual strategy when it leads to many errors (e.g., when a large number of homophonic stimuli are included).

It is possible that the lexical-decision task—which only requires subjects to find a basis on which to discriminate between words and nonwords—may afford decision strategies that cannot be used when the task requires activation of meaning. It is unclear whether the proportion of homophones influences performance on a semantic-decision task, which requires subjects to focus on the meaning of the stimuli. In Van Orden's (1987; Van Orden et al., 1988) experiments and in Experiments 1 and 2, the proportion of homophones was kept low by including large numbers of filler trials. The proportion of homophonic stimuli was 16.4% in Van Orden's first 2 experiments and 10% in his third experiment and in the 2 Van Orden et al. (1988) experiments. In Experiments 1 and 2, the proportion of homophonic stimuli was somewhat higher at 21.4%. In Experiment 3, exactly the same experimental stimuli were used as in Experiments 1 and 2, but the 180 nonhomophonic filler trials per list were replaced by 36 homophone filler trials per list, so that 83.3% of the stimuli were homophones. The spelling controls were the only nonhomophonic stimuli. If subjects can strategically control their use of phonological information, then they should be much less likely to use it in Experiment 3 and thus show little difference in the number of false-positive errors between homophone foils and spelling controls.

Method

Subjects. Twelve McGill University undergraduates were paid $8 each to participate in the study. All were native speakers of English. None had participated in earlier studies.

Stimuli. As in Experiment 2, the category names *living thing* and *object* were used. The 288 experimental stimuli were the same as those used in Experiments 1 and 2 (see Appendix A). The filler words used in the previous experiments were removed from the four lists and replaced by 36 homophone fillers on each list. Half of these were *living things* and half were *objects.* Because it was impossible to find enough homophone words so that the 72 *living thing* and 72 *object* filler homophones needed for the experiment were only used once, 36 of each were found, and every filler appeared on two lists. They were distributed among the four lists such that a list had only 12 homophone fillers in common with any other list.

To summarize, each of the four lists used in the experiment had 108 targets. Half of each were preceded with the category *living thing* and half of each were preceded by the category *object,* and within each category, half of the targets were exemplars and half were not. No more than three trials with the same category name or the same correct response appeared in succession. A practice list of 16 trials, 14 of which had homophone targets, was also developed.

Procedure. The viewing conditions and procedure were exactly the same as in Experiments 1 and 2. Because fewer stimuli were used here than in the previous experiments, each of the four experimental sessions lasted approximately 20 min.

Results

Three sets of analyses were conducted, as in the previous two experiments. The first set examined error rates (false *yes* responses) on homophone foils, spelling controls, and homophone controls; the second set examined correct *no* latencies on these three groups; and the third set compared erroneous *yes* latencies on homophone foils and correct *yes* responses on exemplars. As in Experiment 2, the scores on six words in the homophone control condition (*sun, mail, beach, cellar, prints,* and *buoy*) were not included in the analyses.

False-positive error data. The overall error rate (6.7%) was lower than in Experiment 1 (8.7%) and Experiment 2 (9.8%). The main effect of category–target relation was significant by subjects, $F(2, 22) = 4.58$, $p < .05$, and was marginally significant by items, $F(2, 192) = 2.96$, $p = .054$. Planned comparisons indicated that the overall difference between homophone foils and spelling controls only approached significance by subjects, $F(1, 11) = 4.03$, $p < .07$, and by items, $F(1, 192) = 2.95$, $p < .09$. The difference between homophone foils and homophone controls was significant in both analyses, $F(1, 11) = 7.53$, $p < .02$, by subjects, and $F(1, 192) = 5.52$, $p < .02$, by items. Subjects made 9.6% errors on homophone foils (e.g., object–*break*), 5.7% on spelling controls (e.g., object–*brave*), and 4.9% on homophone controls (e.g., living thing–*break*). In Experiment 2, subjects made 14.2%, 8.8%, and 6.5% errors on these three groups, respectively.

The main effect of target frequency was significant, $F(2, 22) = 10.54$, $p < .001$, by subjects, and $F(2, 192) = 4.44$, $p < .02$, by items. Planned comparisons indicated that subjects made fewer errors on pseudowords (3.2%) than on either low-frequency words (8.3%), $F(1, 11) = 16.84$, $p < .01$, by subjects, and $F(1, 192) = 6.40$, $p < .02$, by items, or high-frequency words (8.7%), $F(1, 11) = 7.04$, $p < .01$, by subjects, and $F(1, 192) = 15.84$, $p < .01$, by items. This is in contrast to the absence of a target frequency effect in Experiment 2. The interaction of category–target relation and stimulus frequency was not significant by subjects, $F(4, 44) = 2.06$, $p > .05$, or by items ($F < 1$). The difference between homophone foils and spelling controls was 1.0% for high-frequency words, 6.0% for low-frequency words, and 4.5% for pseudowords. Planned comparisons indicated that only the difference for pseudowords was significant by subjects, $F(1, 11) = 6.57$, $p < .05$, and none of these differences was significant by items. In Experiment 2, these differences were 0.4%, 8.0%, and 8.0%, respectively. The difference between homophone foils and homophone controls was 0.6% for high-frequency words, 8.9% for low-frequency words, and 4.5% for pseudowords.

The main effect of exemplar frequency was significant by subjects, $F(1, 11) = 5.33$, $p < .05$, but not by items, $F(1, 192) = 1.09$, $p > .05$. The interaction of category–target relation and exemplar frequency was significant by subjects, $F(2, 22) = 10.37$, $p < .001$, and approached significance by items, $F(2, 192) = 2.46$, $p < .09$. Consistent with the results of Experiments 1 and 2 and Van Orden (1987), planned comparisons indicated that homophone foils with low-frequency exemplars produced significantly more errors (8.3%) than spelling controls, $F(1, 11) = 16.82$, $p < .01$, by subjects, and $F(1, 192) = 7.12$, $p < .01$, by items, but the difference for homophone foils with high-frequency exemplars (−0.6%) was not significant. The difference between homophone foils and homophone controls was 8.7% for those with low-frequency exemplars and 0.4% for those with high-frequency exemplars. In Experiment 2, significant effects were also observed only for homophone foils with low-frequency exemplars.

The triple interaction was not significant by subjects, $F(4, 44) = 2.11$, $p > .05$, or by items ($F < 1$). Planned comparisons performed to examine the difference in error rate between homophone foils (e.g., object–*break*) and spelling controls (e.g., object–*brave*) revealed that the difference was significant for low-frequency words with low-frequency exemplars (14.6%), $F(1, 11) = 16.25$, $p < .01$, by subjects, and $F(1, 192) = 5.95$, $p < .02$, by items, and was significant by subjects for pseudowords with low-frequency exemplars (8.3%), $F(1, 11) = 7.07$, $p < .05$, but not by items, $F(1, 192) = 2.62$, $p > .05$. These were the two groups that produced the largest differences in Experiment 2. In Experiment 2, the item analysis for pseudowords approached significance ($p < .07$); here, it was not significant ($p = .11$). None of the other differences approached significance. The differences in mean error rates between homophone foils and spelling controls for the six groups are presented in Figure 4.

Planned comparisons between homophone foils (e.g., object-*break*) and homophone controls (e.g., living thing-*break*) indicated that significantly more false-positive errors were made on homophone foils than on homophone controls for low-frequency words with low-frequency exemplars (16.0%), $F(1, 11) = 24.01$, $p < .001$, by subjects, and $F(1, 192) = 7.67$, $p < .01$, by items, and the difference was significant by subjects for pseudowords with low-frequency exemplars (9.0%), $F(1, 11) = 7.07$, $p < .05$, and approached significance by items, $F(1, 192) = 3.13$, $p < .08$. These were the two groups that produced significant differences in Experiment 2. The differences for high-frequency foils with low-frequency exemplars (1.8%) and high-frequency foils (−0.6%), low-frequency foils (1.8%), and pseudohomophone foils (0%) with high-frequency exemplars were not significant.

In sum, the overall size of the difference between homophone foils and spelling controls was smaller (by 1.5%) than in Experiment 2 and only approached significance here. The differences were largest in the same two groups as in Experiment 2. The overall size of the difference between homophone foils and homophone controls was also smaller (by 3%) than in Experiment 2 but remained significant. The difference was also largest in the same two groups as in Experiment 2. In contrast to the absence of a target frequency effect in Experiment 2, here fewer errors were made on pseudowords than on high- or low-frequency words.

No *reaction time data.* As in Experiments 1 and 2, a subject's response latency for an item was only included in the analyses if the subject responded correctly to it and the other three members of the stimulus quadruple (e.g., object-*brake*, object-*break*, object-*brave*, and living thing-*break*). This criterion was met by 73.7% of responses. Another 19.4% of responses were correct but were discarded because an error was made on another member of the quadruple. Essentially the same results were found when all correct reaction times were included. Five percent of response times were greater than 1,500 ms and were replaced with times of 1,500 ms. Subjects responded more slowly in Experiment 3 than in Experiment 2 by an average of 124 ms.

The main effect of category-target relation in the *no* latency data was not significant, $F(2, 22) = 2.94$, $p > .05$, by subjects ($F < 1$, by items). Decision latencies were 919 ms for homophone foils, 915 ms for spelling controls, and 896 ms for homophone controls. In Experiment 2, the difference between homophone foils and spelling controls was 11 ms, and the difference between homophone foils and homophone controls was 34 ms.

The main effect of target frequency was significant by subjects, $F(2, 22) = 13.88$, $p < .001$, and by items, $F(2, 192) = 13.73$, $p < .001$. Subjects responded more quickly to pseudowords (865 ms) than to high-frequency words (935 ms) and low-frequency words (930 ms). This is in contrast to the results of Experiment 2 in which high-frequency targets were responded to faster than low-frequency words and pseudowords.

No other differences were found in Experiment 2, and here there were two further effects that were significant in the subject analysis only. The interaction of category-target relation and target frequency was significant by subjects, $F(4, 44) = 7.72$, $p < .001$, but not by items, $F(4, 192) = 1.50$, $p > .05$, and the interaction of category-target relation and exemplar frequency was significant by subjects, $F(2, 22) = 4.89$, $p < .02$, but not by items ($F < 1$). Neither the main effect of exemplar frequency nor the triple interaction was significant by subjects or items.

Yes *response latencies.* An item was included in these analyses only if a subject made both a false-positive error on

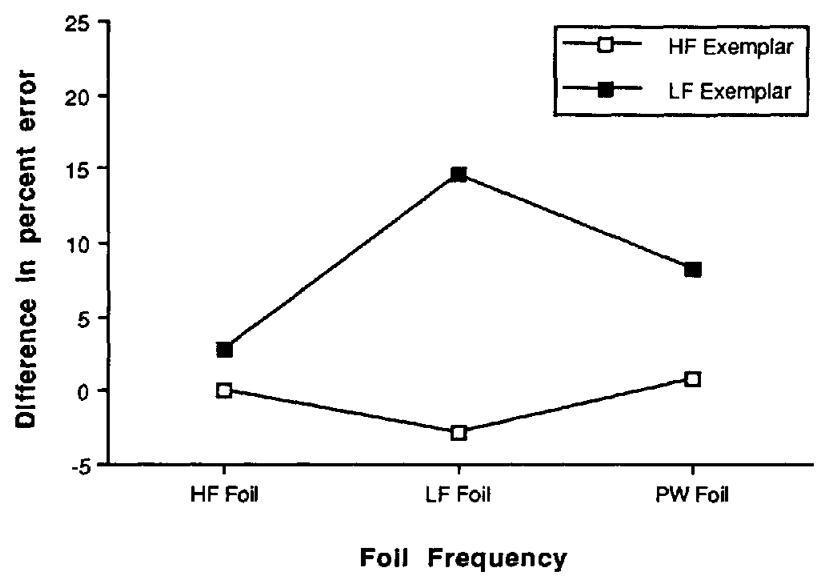

Figure 4. The difference in mean percentage errors between homophone foils and spelling controls in Experiment 3. (HF = high frequency; LF = low frequency; and PW = pseudoword.)

the homophone foil and correctly accepted the matched exemplar. Seven out of 82 false-positive latencies on homophone foils were not included because errors were made on the corresponding exemplars. Subjects responded *yes* 12 ms faster to exemplars than to homophone foils, but the difference was not significant by subjects ($t < 1$) or by items, $t(62) = 1.28$, $p > .05$. The difference was 94 ms for low-frequency foils with low-frequency exemplars and their matched exemplars, but this difference was not significant, probably because only 24 false-positive errors were made on words in this group. Fewer were made in each of the other groups.

Discussion

The subjects in Experiment 3 responded more cautiously than subjects in Experiment 2. They made 3.1% fewer errors than subjects in Experiment 2, and their decision latencies were 124 ms slower on average.

The evidence for phonological activation of meaning was weaker than in Experiment 2. The overall difference between homophones and spelling controls was 1.5% smaller than in Experiment 2 and only approached significance. The difference for low-frequency words was 2% smaller and the difference for pseudowords was 3.5% smaller, and both were no longer significant. However, low-frequency foils with low-frequency exemplars still produced significantly more errors than spelling controls, and in fact, the numerical difference was the same as in Experiment 2 (14.6%). The difference between pseudohomophone foils with low-frequency exemplars and their spelling controls dropped by 2.1% to 8.3%, and although the difference was still significant by subjects, it no longer approached significance by items.

Van Orden et al. (1988) argued that evidence that the phonological information responsible for these effects is computed before, rather than after, the activation of meaning is the finding of a similar-sized homophone foil effect for words and pseudowords. One possible interpretation of the larger effect for low-frequency words here is that it may reflect, at least in part, the activation of phonological information after the activation of meaning. However, several aspects of the data suggest instead that it was the pseudowords that were processed rather differently in Experiment 3. Although the overall drop in error rate from Experiment 2 to Experiment 3 was 3.1%, the drop was 5.9% in errors made on pseudowords and only 0.8% for low-frequency words. In Experiment 3, subjects produced significantly fewer errors and faster latencies on pseudowords than on either high- or low-frequency words. In contrast, in Experiment 2, there was no effect of frequency on errors, and in the latency data, high-frequency words were responded to most quickly, low-frequency words had longer latencies, and pseudowords had the longest latencies. These observations suggest that the manner in which the task was performed changed in a way that facilitated the detection of pseudowords. One possibility is that under conditions in which decisions are difficult, subjects may also use orthographic familiarity information to detect pseudowords. This additional information would be useful only for pseudowords because the correct response for a pseudoword is always *no*, but the correct response for a word depends on its meaning.

The results of Experiment 3 were again consistent with Van Orden's account of a spelling-check procedure, because subjects made more errors on homophone foils when they had low-frequency exemplars than when they had high-frequency exemplars. The prolonged *no* decision latencies, particularly for words, suggest that subjects were performing this check more carefully than in Experiment 2 because of the higher proportion of homophones. The failure to observe a significant difference between *yes* latencies on exemplars and false-positive latencies on foils is also consistent with the observations by Van Orden et al. (1988). They claimed that this was evidence that orthographic information does not contribute to *yes* responses. However, there was a 94-ms difference between low-frequency foils and matched low-frequency exemplars that was very likely not significant because it was based on only 24 responses. This observation suggests that it would be premature to draw the conclusion that there is no influence of orthography from the present data. Further evidence from a much larger number of false-positive errors is needed.

In summary, the results of Experiment 3 suggest that when a high proportion of the stimuli are homophones, the semantic-decision task is performed more slowly and performance on pseudowords is better relative to words, but the pattern of responses on words changes very little. These results suggest that the use of phonological information is not strategically controlled, contrary to previous claims (Coltheart, 1978; Davelaar et al., 1978; Hawkins et al., 1976; McQuade, 1981). Coltheart argued that the visual route is usually faster than the phonological route. If including a high proportion of homophones in the experiment forces subjects to abandon the phonological route, Coltheart's view suggests that latencies should have become faster, not slower. In addition, if subjects were using a visual recognition strategy, no effect of homophony would be expected; however, this effect was observed for low-frequency foils with low-frequency exemplars. The results of Experiment 3 suggest that the effect of including many homophones was not to change the type of representation subjects used to activate meaning, but rather to extend the orthographic checking process. This, perhaps in combination with an orthographic familiarity strategy, made pseudowords more likely to be detected.

Previous evidence for the ability to avoid the use of phonology came from tasks that do not require the activation of meaning. Hawkins et al. (1976) used a tachistoscopic task that can be performed on the basis of orthographic information. Davelaar et al. (1978) and McQuade (1981) used a lexical-decision task, which—as discussed in the introduction—may be performed on the basis of orthographic familiarity (Balota & Chumbly, 1984; Besner et al., 1984; Gordon, 1983; Seidenberg & McClelland, 1989; Waters & Seidenberg, 1985). Thus, the absence of phonological effects in these studies when a high proportion of homophonic stimuli were included may not reflect the absence of phonological activation of meaning, but rather may reflect a switch to a superficial orthographically based decision strategy. No firm conclusion regarding the use of phonology to activate meaning can be made, then, from these studies. The category-decision task, on the other hand, requires that subjects consult the meanings of words and thus is more suited to providing evidence about the representations used to activate meaning. The results of Experiment 3 suggest that the use of phonological information is not an optional strategy for low-frequency words.

Experiment 4

The finding in Experiment 3 that subjects did not avoid the use of phonology when the experiment contained a high proportion of homophones was surprising given previous claims that subjects can strategically control their use of phonological information. The conclusion that phonological activation of meaning was occurring was based on the observation that subjects produced more false-positive errors on low-frequency foils with low-frequency exemplars than on spelling controls. Experiment 4 was designed to replicate this result with a larger number of stimuli. Only 12 words were included in each group in Experiments 1–3, primarily because of the difficulty in finding high-frequency homophones. More homophones are available if only low-frequency pairs are required. Fourteen new homophones were added in Experiment 4. Two words, *flea* and *pole*, which had produced more errors than other homophone foils in Experiments 1–3, were not included to ensure that these words were not unduly skewing the results. The experiment again used a large proportion of homophone filler trials to determine whether there is phonological activation of meaning even when the stimuli would seem to discourage this type of processing.

In addition, Experiment 4 included a spelling similarity manipulation designed to determine whether more false-positive errors are made on homophone foils, such as *alter*, that are very similar in spelling to their exemplars (*altar*) than on homophone foils, such as *slay*, that are much less similar to their exemplars (*sleigh*). Van Orden (1987) argued that if a spelling check is performed on phonologically activated candidates, then more errors should be made when exemplars and foils are similarly spelled. A phonological impostor (the exemplar) should be more likely to slip by the spelling check if its spelling is similar to the target foil than if its spelling is dissimilar. Van Orden found a larger difference between homophone foils and spelling controls when exemplars were spelled similarly to their foils. This effect should be especially strong in Experiment 4 if the consequence of including a large proportion of homophones is to force subjects to perform this check more carefully.

In summary, if the results of Experiment 3 replicate, and phonological activation of meaning cannot be prevented, a significant difference between homophone-foil and spelling-control errors should be found. If a spelling check is performed, more errors should be made on foils that are similar to their exemplars than on dissimilar foils. The homophone control condition was not included in Experiment 4 because both of the control conditions in all three previous experiments yielded very similar results. These results are sufficient to establish that the higher error rate for the homophone foils does not occur simply because they are homophones.

Method

Subjects. Twelve McGill University undergraduates were paid $5 each to participate in the study. All were native speakers of English; none had participated in earlier studies.

Stimuli. As in Experiments 2 and 3, the category names were *living thing* and *object.* There were 72 experimental words (these are presented in Appendix B) and 198 filler words. Twenty-four of the experimental words were homophone exemplars, 24 were the other member of the homophone pair and served as homophone foils, and 24 were spelling controls. All were low-frequency words. Twelve of the exemplar–foil–spelling control triples had foils and spelling controls that were spelled similarly to their exemplars (e.g., *altar–alter–ajar*). The mean similarity of the foils to the exemplars (using Weber's, 1970, measure) was .72, and the mean similarity of the spelling controls to the exemplars was .76. The remaining 12 triples had foils and spelling controls that had spellings that were not similar to their exemplars (e.g., *sleigh–slay–slam*). The mean similarity of the foils to the exemplars was .52, and the mean similarity of the spelling controls to the exemplars was .50. (Theoretically, values on this measure range from 0 to 1, but few homophone pairs have similarities less than .40 or greater than .80.) Exemplars, foils, and spelling controls were matched for frequency and length. The mean frequencies, respectively, were 4.3, 4.6, and 6.3, for the similarly spelled triples, and 5.3, 5.8, and 5.9, for the dissimilarly spelled triples. Half of the exemplars in each group were members of the category *living thing* and half were members of the category *object.* Only three lists were needed because the homophone-foil condition was not included in this experiment. Each member of a stimulus triple (exemplar, foil, and spelling control) was placed on a different list, with the result that 24 experimental words, 8 of each category–target relation, appeared on each list.

Sixty-six filler words also appeared on each list, and 56 of these were homophones. Sixteen of the homophone fillers were *living things*, 16 were *objects*, and 24 were not exemplars of their categories. Five of the nonhomophone fillers were members of their categories and five were not members. Thus, of the 90 words on each list, 45 were preceded by the category *living thing* and 45 were preceded by the category *object*, and within each category, half of the words were exemplars and half were not. Each list was composed of 80% homophones and 20% nonhomophones. No word appeared more than once in the experiment, and no pseudowords were included. Sixteen other words were chosen for a practice list. Half of the words were preceded by the category *living thing* and half were preceded by the category *object*, and half of each were members of their categories.

Procedure. The viewing conditions and procedure were exactly the same as in Experiments 1–3 except that the experiment required only three 15-min sessions.

Results

One word triple in the high-similarity group (object: *cord, chord, cod*) was removed from the analysis because 10 of the 12 subjects made an error on the spelling control (*cod*), and the 2 others had latencies of longer than 1,500 ms. This item was ambiguous because it is a *living thing*, but as a prepared food it could also be considered an *object.*

There were two factors in the analyses of the false-positive error and the *no* latency data, orthographic similarity (similar vs. dissimilar) and category–target relation (foil vs. spelling control). The data were analyzed as in the previous experiments.

False-positive error data. The differences in percentage errors between homophone foils and spelling controls are

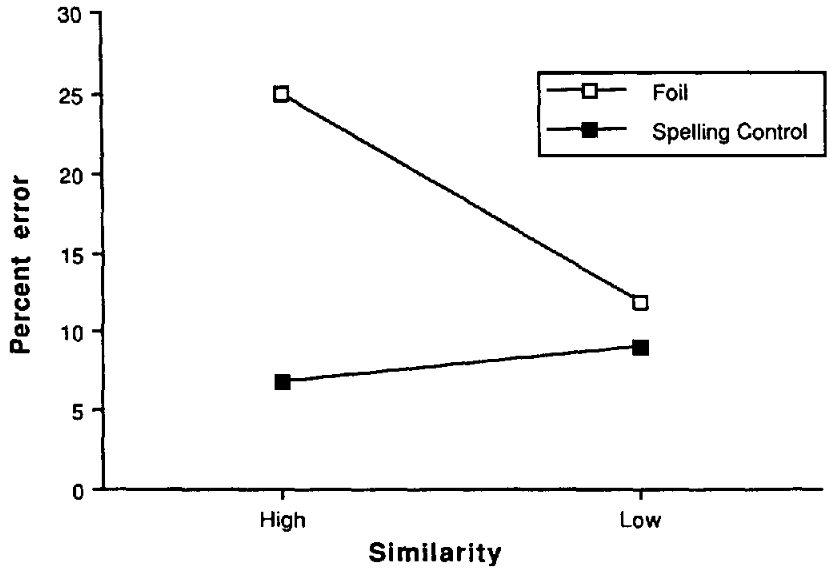

Figure 5. Mean percentage errors for foils and spelling controls in Experiment 4 as a function of their similarity to corresponding exemplars.

presented in Figure 5. There was a main effect of category–target relation that was significant by subjects, $F(1, 11) = 9.46$, $p < .01$, and was marginally significant by items, $F(1, 42) = 3.88$, $p = .056$. Subjects made more errors on homophone foils (18.4%) than on spelling controls (7.9%). The interaction of category–target relation and similarity was significant by subjects, $F(1, 11) = 6.95$, $p < .05$, but not by items, $F(1, 42) = 1.88$, $p > .05$. Simple main effects revealed a significant difference between homophone foils and spelling controls for foils that are spelled similarly to their exemplars (18.2%), $F(1, 11) = 13.12$, $p < .01$, by subjects, and $F(1, 42) = 5.35$, $p < .05$, by items, but not for dissimilarly spelled words (2.8%; both Fs < 1).

No *reaction time data.* A subject's response latency was only included in the analyses if the subject responded correctly to it and the other two members of the stimulus triple. This criterion was met by 66.7% of responses. Another 19.0% of responses were correct but were discarded because the subjects made an error on another member of the stimulus triple. A total of 5.6% of response times were greater than 1,500 ms and were replaced with times of 1,500 ms.

Neither the main effect category–target relation nor the interaction of category–target relation and similarity was significant (all Fs < 1). Subjects took 20 ms longer to make decisions about homophone foils than spelling controls when they were spelled similarly and 35 ms longer when they were spelled dissimilarly.

Yes *response latencies.* An item was included in these analyses only if a subject made both a false-positive error on the homophone foil and correctly accepted the corresponding exemplar. One out of 40 false-positive latencies on homophone foils was not included because the subject failed to respond correctly to the corresponding exemplar. There was no difference between *yes* latencies for homophone foils (898 ms) and exemplars (896 ms; both ts < 1). The difference for the similarly spelled group alone was not significant either (both ts < 1).

Discussion

Low-frequency homophone foils with low-frequency exemplars produced more errors than spelling controls in Experiment 4. Because additional words were included in this experiment, it suggests that the effect found for this group in Experiment 3 was not specific to the small set of words used. This is further evidence for the claim that the use of phonological information cannot be strategically controlled in a task that requires subjects to consult the meanings of words. If subjects had been able to prevent phonological activation of meaning, there should have been no difference between homophone foils and spelling controls. Experiment 4 also demonstrated that the difference in error rates on homophone foils and spelling controls is larger when the foil and exemplar are similarly spelled. This suggests that subjects were performing a spelling check on phonologically activated candidates and that the spelling check was less likely to catch a phonological impostor if its spelling was similar to the target homophone foil. Thus, Experiments 1–4 have demonstrated that homophone foils do not in general produce more errors than spelling controls. Rather, the effect is limited to low-frequency homophone foils that are spelled similarly to their low-frequency exemplars.

Discussion of Experiments 1–4

According to Van Orden (1987), a visually presented letter string activates its phonological representation, which in turn activates a candidate set of meanings. The most active of the candidates is submitted to a spelling check, in which its spelling is compared with that of the target. False candidates are more likely to pass the spelling check if their spelling is unfamiliar, if they are spelled similarly to the target, or both. This proposal predicts that more false-positive errors should be made on homophone foils than on spelling controls when foils have low-frequency exemplars or when exemplars are spelled similarly to the target foils. The results of Experiments 1–4 support these predictions. The observation that the use of phonological information is not strategically controlled by the subject is also consistent with Van Orden's view.

However, the present results suggest that Van Orden et al. (1990) and especially Van Orden (1987; Van Orden et al., 1988) overstated the extent to which phonological information contributes to the activation of meaning. No evidence for phonological activation of the meanings of high-frequency words was found in Experiment 2 or 3 in which priming from the category name was reduced. Van Orden (1987) assumed that the exemplar would always be more highly activated than the foil and would be submitted to the spelling check first, because it is consistent with both the phonological representation and the category name. In this view, high-frequency foils with low-frequency exemplars should have produced more errors than spelling controls, and they did not. However, the present results can be accommodated if it is assumed that the highest frequency member of the homophone pair is submitted to the spelling check first and not necessarily the exemplar. This might occur when information available from category names is reduced through the use of broad categories. In the case of high-frequency foils with low-frequency exemplars, the foil would be submitted to the spelling check first, and because its spelling is familiar, it should easily pass. Thus, there should be no more errors and no longer latencies than for spelling controls. In the case of high-frequency foils with high-frequency exemplars, the exemplar would be submitted to the spelling check first when it was higher in frequency, but because its spelling is familiar, the spelling check should easily detect that it is not the target presented. The foil would be submitted next, and again an error should not be made because its spelling is familiar. Thus, for this group, the theory would predict longer latencies for foils than for spelling controls and no effect in the error data. The effect in the latency data would be weak because only the exemplars that are higher frequency than the foil would be chosen for the spelling check first. In this modified view, then, differences between high-frequency homophone foils and spelling controls would not likely be observed because the spelling check would make few errors on these words. Empirically, it would be difficult to distinguish this explanation from the view that phonological mediation does not occur for high-frequency words. The only case in which this modified view predicts more errors on foils than on spelling controls is when the spelling of both the foil and the target are not familiar. This is what was observed in Experiments 2 and 3.

A consequence of the above analysis is that a comparison of false-positive error rates on homophone foils and spelling controls is not capable of unambiguously determining whether the meanings of high-frequency foils are phonologically activated and the spelling check is efficient or whether they are activated on a visual basis. Some evidence might be obtained from the *no* latency data, although a problem with making inferences from these data was pointed out earlier. *No* latencies may include postrecognition semantic processing that may mask small effects on recognition itself. There was no difference between high-frequency homophone foils with high-frequency exemplars and their spelling controls in the *no* latency data in either Experiment 2 or 3.

Van Orden et al. (1988) argued that the observation of similar *yes* latencies on exemplars and homophone foils indicated that the exemplars did not benefit from having orthographic information consistent with the *yes* response. In Experiments 2–4, there was no significant difference between foil and exemplar *yes* latencies; however, it was argued that strong conclusions could not be drawn from these data because incorrect *yes* latencies may arise from an earlier point in processing than correct *yes* latencies and because the analyses were based on too few items. The analyses of Van Orden et al. (1988) were also based on a small number of responses. In their Experiment 1, the analysis of word latencies was based on an average of 2.1 scores per subject, and in their Experiment 2, it was based on an average of 3.3 scores per subject. Thus, it is unclear from these data whether there is orthographic activation of meaning. However, the finding that specific category names can prime exemplars suggests that some orthographic information contributed to the activation of word meanings in Van Orden's (1987, Experiment 2) tachistoscopic experiment. The false-positive error rate should have been much higher in his experiment if no orthographic information were available.

In summary, evidence from the present experiments that the frequency of the foil is important contradicts Van Orden's (1987; Van Orden et al., 1990) position. However, if Van Orden's (1987) assumption that the exemplar is always submitted to the spelling check first is dropped, then there is some support for a verification view, because the results of the present studies, particularly the finding of fewer errors on foils with high-frequency exemplars than on foils with low-frequency exemplars, suggest that subjects perform a spelling check. Further data on high-frequency words are required to determine whether their meanings are phonologically activated. Also, more evidence is required to determine whether orthographic information directly activates word meanings. If such evidence were found, it would contradict Van Orden's (1987; Van Orden et al., 1988) hypothesis that meaning is activated exclusively through phonology.

The time-course dual-route view (Seidenberg, 1985a, 1985b), in contrast, assumes that skilled readers quickly activate the meanings of high-frequency words on a visual basis and therefore correctly predicts no difference between high-frequency homophone foils and spelling controls. For low-frequency words, activation from orthography to meaning is assumed to be slower, and so phonological information can influence the activation of the semantic representation (Seidenberg, 1985a, 1985b; Seidenberg & McClelland, 1989). This view correctly predicts a difference between homophone foils and spelling controls for low-frequency words. Because, in this view, the meanings of low-frequency words are activated from orthography and phonology, it can also provide an account for the finding that more errors are made on homophone foils that are spelled similarly to their exemplars. When low-frequency foils are presented (e.g., *alter*, *slay*), the meanings associated with similarly spelled exemplars (e.g., *altar*) will receive activation from both orthographic (e.g., *A*, *L*, *T*, *R*) and phonological codes, but dissimilarly spelled exemplars (e.g., *sleigh*) will receive activation primarily from phonology. Similarly spelled exemplars will thus be better able to compete with the foil. Finally, this theory provides an explanation for Van Orden's (1987, Experiment 2) tachistoscopic experiment results. He found fewer errors on homophone foils than might be expected given that the exemplar was primed. Activation of the foil's meaning from orthography could have provided some information that served to reduce errors.

This dual-route theory does, however, have difficulty accounting for the larger number of errors made on foils with low-frequency exemplars than those with high-frequency exemplars. High-frequency exemplars should be activated more strongly than low-frequency exemplars and thus should be better able to compete with the foil. To account for effects of exemplar frequency, a spelling check could be added that is used when two or more words are highly activated. This could involve waiting until more evidence accumulates from the orthographic route or examining the orthographic code associated with each meaning to see if it matches the input orthographic code. In each case, however, a delay in processing would be expected for homophones relative to spelling controls when they are low in frequency, and no such effect was observed. Finally, the finding that subjects cannot strategically control the use of phonological information contradicts Coltheart's (1978) view, but is consistent with Seidenberg and McClelland's (1989) model and with demonstrations by a number of researchers that phonology is automatically activated in visual word recognition (e.g., Bakan & Alperson, 1967; Dalrymple-Alford, 1972; Humphreys, Evett, & Taylor, 1982; Perfetti, Bell, & Delaney, 1988; Tanenhaus, Flanigan, & Seidenberg, 1980).

In summary, dual-route theories as proposed by Coltheart (1978), Seidenberg (1985a, 1985b), Seidenberg and McClelland (1989), and others do not assume that a spelling check normally occurs in word recognition. In the case of homophones, the parallel activation of meaning from orthography would normally be sufficient to establish which meaning is correct relative to the category. However, the finding of more errors on foils with low-frequency exemplars than on foils with high-frequency exemplars can only be explained by assuming that a spelling check took place.

The failure to find significant false-positive effects for higher frequency words in Experiments 2 and 3 would seem to contradict Van Orden's claim that meaning is activated exclusively (Van Orden, 1987; Van Orden et al., 1988) or at least predominantly (Van Orden et al., 1990) by phonological representations. However, we have suggested that the absence of these effects is somewhat ambiguous: They could result because meaning is activated directly form orthography or because the spelling-check mechanism does not err on these words. Experiments 5 and 6 provide another way of examining the role of phonology in the processing of higher frequency words. In addition, these studies examine whether the spelling check is a necessary component of word recognition, as Experiments 1–4 and the verification theory suggest, or whether it is an experiment-specific strategy, as dual-route theory suggests.

Experiment 5

The results of Experiments 1–4 suggest that the meanings of low-frequency words are phonologically activated and that a spelling check is performed on activated candidates. However, it is possible that the spelling check is a strategy used by subjects when the experiment contains homophone foil trials. Orthographic information from the foil target may give its semantic representation a small advantage over that of the exemplar, but when subjects become aware that some targets sound like an exemplar of the category but are not exemplars, they may check their spellings carefully to prevent errors. Van Orden (1987) acknowledged that his findings could not determine whether the spelling check is always performed or is only performed when the experiment contains homophone foils. If a spelling check is performed only when the experiment contains homophone foils, this would imply that there is usually enough orthographic information available to distinguish between members of a homophone pair. That is, there is activation of meaning from orthography, the direct route.

One aim of Experiment 5 was to examine whether subjects use a spelling check when no homophone-foil trials are included in the experiment. Instead of examining performance on homophone foils, responses on homophone exemplars (e.g., living thing-*flea*) were examined. The goal of the experiment was to determine whether homophony has an impact on correct *yes* responses. The verification view still predicts that the spelling check will occur, and if the assumption that the exemplar is always submitted to the spelling check first is dropped, then subjects should make fewer errors on exemplars when the other member of the homophone pair is high in frequency. When the spelling of the other member is familiar, the spelling check should be better able to detect that it does not match the exemplar target than when its spelling is unfamiliar. The dual-route view, on the other hand, predicts that subjects should be more likely to make an error on a homophone exemplar if the other member is high in frequency. The meaning of a high-frequency foil would be activated more strongly than that of a low-frequency foil and thus would be better able to compete with the meaning of the target exemplar. (We continue to use the term *foil* to refer to the member of the homophone pair that is not an exemplar of the category. Note, however, that the exemplars, not the foils, were presented as targets in these experiments.)

To be able to examine effects of the frequency of the foil in skilled readers who make few errors on target words that are correct exemplars of the category, Experiment 5 compared performance on homophone exemplars with performance on these exemplars when the foil was made more salient by priming it with a semantically related word. That is, in one condition, exemplars were preceded by a word semantically related to the foil (e.g., the exemplar *brake* was preceded by *shatter*, related to the foil *break*), and in the other, they were preceded by an unrelated word (e.g., *bold–brake*). If a spelling check is carried out, priming the foil should make it more likely to be checked first, and foils with unfamiliar spellings are more likely to erroneously slip by. Thus, more errors should occur on exemplars with low-frequency foils. In contrast, according to dual-route theory, increasing the activation of the foil should cause subjects to make more errors or produce longer decision latencies on homophone exemplars with high-frequency foils.

The second aim of Experiment 5 was to see whether effects of phonology could be observed for high-frequency words using a more sensitive measure—correct *yes* latencies—than in the previous experiments. In the previous experiments, evidence for the use of phonology came primarily from error responses, and subjects did not make many errors on high-frequency words. It was argued that the failure to find effects for high-frequency words could occur either because access to their meanings was phonologically activated but the spellings were well known so the spelling check did not err or because their meanings were activated directly on the basis of visual information. There were no clear effects of homophony for high-frequency words in the *no* latency data, but time to decide that a word is not an exemplar of a category may include considerable processing after the meaning of the word had been obtained. Thus, correct *yes* latencies to homophone exemplars should be a more sensitive measure of effects of phonology. If both the correct and foil meanings of high-frequency homophones are activated to some extent, then increasing the activation of the foil meaning through priming should cause subjects to produce longer decision latencies or make more errors on the homophone exemplars. If subjects quickly determine the meanings of homophones before semantic representations receive activation from phonological representations, then there should be no effect of the priming manipulation.

Underwood and Thwaites (1982) conducted a lexical-decision experiment that also made use of pairs of words in which one was a homophone (e.g., *waist*) and the other was a word semantically related to the other member of the homophone pair (e.g., *rubbish*). The two words were presented simultaneously; the semantically related words appeared in the same location as that of the fixation point, and the homophones were presented to the right of fixation and were pattern-masked. The subjects' task was to decide whether the stimulus that appeared centrally (e.g., *rubbish*) was a word. Underwood and Thwaites found that response latencies to central words were slowed when the word in the periphery was a homophone (e.g., *waist*) whose other member was related to the target compared with a condition when the word in the periphery was an unrelated homophone. They attributed this effect to the use of phonological information in decision processes because the peripherally presented word would usually be processed after the centrally presented target word. In Experiment 5, the word semantically related to the foil was presented before the homophone and thus could influence early processing of the homophone.

Preliminary Study

Before the main experiment was conducted, a preliminary study was run. This was needed to establish that the semantically related words to be used in the main experiment actually did prime the foil (e.g., that *shatter* primes *break*). In the preliminary study, two semantically related words were chosen for each foil (e.g., *shatter*, *fracture*). Foils were presented to subjects three times, in separate sessions, once preceded by each semantically related word and once preceded by an unrelated word. The semantically related word that caused the greatest decrease in response latency for each foil was chosen as the prime for the main experiment. An analysis was then done to ensure that these primes did indeed result in significantly faster response latencies on foils than when foils were preceded by an unrelated word.

Neely, Keefe, and Ross (1989) argued that semantic priming effects could be due either to automatic spreading activation or to strategies, such as the generation of expectancies and postlexical semantic matching, depending on how the task is set up. To reduce the likelihood that priming effects were due to subject strategies, a continuous semantic-decision task was used. The subjects were told the name of a category at the beginning of a block of trials. Words were presented one at a time, and subjects made decisions to each item. Primes appeared as a normal trial before the target trial. Subjects were not told that some successive trials were related. Furthermore, the proportion of related trials was kept low by including filler words unrelated to preceding words (see Neely et al., 1989; Seidenberg, Waters, Sanders, & Langer, 1984; Tweedy, Lapinsky, & Schvaneveldt, 1977), and none of the primes were high associates of the targets (see Fischler, 1977).

Method

Subjects. Eighteen McGill University undergraduates were paid $8 each to participate in the study. Another 15 subjects volunteered to fill out a questionnaire. All were native speakers of English.

Stimuli. The target stimuli were the homophone word foils that were used in Experiments 1–3 (e.g., object–*break*). The pseudohomophone foils (e.g, object–*shews*) were not included because in the main study the exemplars were presented, and their exemplars are not homophones (e.g., *shoes*). For each of the 48 foils, two semantically related words were chosen to serve as primes (e.g., *shatter* and *fracture* for *break*). Three experimental lists were created with each *living thing* foil appearing once on each list, and three lists were created with each *object* foil appearing once on each list. Across the three lists, foils appeared once preceded by an unrelated word, once preceded by one of the semantically related words, and once preceded by the second semantically related word. These trials were distributed among the lists such that on 8 of the 24 experimental trials on each list, the foil was preceded by an unrelated word, and on 16 of the trials, the foil was preceded by a semantically related word. An additional 112 fillers were included on each list; 80 were exemplars of the category and 32 were not exemplars. The fillers reduced the percentage of related trials to 10% of the trials on a list. Thus, each of the six lists contained 80 *yes* trials and 80 *no* trials. The fillers were placed randomly on each list, although not between prime–target pairs. The category names did not appear on the lists because the same decision was to be made about all words on a given list. Two practice lists were created, one for the *living thing* category and one for the *object* category. Half of the items on each list were members of the category and half were not.

A questionnaire was created that listed each of the 96 semantically related prime words with a blank line beside each one. Instructions on the top of the page asked subjects to write down the first related word that came to mind for each of the words on the list. These data were collected to ensure that the targets were not high associates of the primes (e.g., *bread–butter*).

Procedure. The questionnaire was distributed to 15 subjects who were asked to fill in the first related word that came to mind for each word on the list. They were instructed not to spend too much time on any item. The 18 other subjects completed three experimental sessions, each lasting 30 min. The sessions were separated by at least 1 week. In a session, the subjects were told the name of a category and saw the practice trials followed by a list of 160 trials for that category and then were told the name of the other category and saw the practice trials and a list of 160 trials for that category. Subjects were given feedback on each practice trial to ensure that they understood the task. Subjects saw each of the six lists once. The order of presentation of the experimental lists was counterbalanced across subjects such that each list was presented in each of the six list positions three times. Words were presented on the screen one at a time and remained until the subject responded. Subjects were not told that words on some trials were semantically related to words on the subsequent trial. The intertrial interval was 1 s. Other aspects of the procedure were the same as in the previous experiments.

Results and Discussion

The 2.8% of response times that were greater than 1,500 ms were replaced with times of 1,500 ms. Item means were then calculated for foil targets (e.g., *break*) preceded by each of their semantically related primes (e.g., *shatter, fracture*). The prime that was associated with the shortest response time on the succeeding foil was chosen. These are presented in Appendix C. The questionnaires were examined to see how often subjects produced the foil target given the chosen prime word. Targets were produced on 32 occasions out of a possible 720 (4.4%). In one case (*encounter*), 6 out of 15 subjects produced the foil (*meet*); in three cases, 4 out of 15 subjects produced the foil; and in the nine other cases, 3 or fewer of the subjects produced the foil. Thus, the chosen semantically related prime words are not high associates of the foil targets.

Responses on foil targets when they were preceded by the unrelated word and when they were preceded by the chosen prime word were submitted to analyses of variance. There were three factors in the analyses: prime condition (primed vs. unprimed), foil frequency (high vs. low), and exemplar frequency (high vs. low). They were all treated as within factors in the analyses by subjects. In the analyses by items, prime condition was treated as a within factor (because exactly the same words were used in both conditions), and the other two factors were treated as between factors. Percentage errors were arcsine transformed before analysis.

There was a significant effect of prime condition in the error data, $F(1, 17) = 13.04$, $p < .01$, by subjects, and $F(1, 44) = 10.65$, $p < .01$, by items. Subjects produced fewer errors on foils when they were preceded by a semantically related prime (6.6%) than when they were preceded by an unrelated prime (10.8%). The interaction between prime condition and frequency of the foil was not significant, $F(1, 17) = 1.83$, $p > .05$, by subjects, and $F(1, 44) = 2.58$, $p > .05$, by items, nor was the interaction between prime condition and frequency of the exemplar (both Fs < 1) or the triple interaction between prime condition, frequency of the foil, and frequency of the exemplar, $F(1, 18) = 1.71$, $p > .05$, by subjects, and $F(1, 44) = 1.94$, $p > .05$, by items. The size of the priming effect was 3.7% for high-frequency foil with high-frequency exemplars, 3.7% for high-frequency foils with low-frequency exemplars, 1.0% for low-frequency foils with high-frequency exemplars, and 8.3% for low-frequency foils with low-frequency exemplars.

There was a significant main effect of prime condition in the decision latency data, $F(1, 17) = 18.89$, $p < .001$, by subjects, and $F(1, 88) = 7.30$, $p < .01$, by items. Subjects produced faster responses on foils when they were preceded by a semantically related prime (709 ms) than when they were preceded by an unrelated prime (756 ms). There was no interaction between prime condition and frequency of the foil, or between prime condition and frequency of the exemplar, and no significant triple interaction (all Fs < 1). The size of the priming effect was 48 ms for high-frequency foils with high-frequency exemplars, 45 ms for high-frequency foils with low-frequency exemplars, 48 ms for low-frequency foils with high-frequency exemplars, and 45 ms for low-frequency foils with low-frequency exemplars.

The preliminary study served to identify a semantically related word (e.g., *shatter*) for each homophone foil (e.g., object–*break*) and demonstrated that these words do in fact prime the foils in a continuous semantic-decision task. Insofar as the prime and target trials were not presented as pairs, a low proportion of the trials were related, and the targets were not high associates of the primes, it is very likely that these effects were not due to subjects anticipating the targets or engaging in postlexical comparisons of prime and target.

The words found to prime the foils in the preliminary study were then used as primes for the exemplars in a similar continuous semantic-decision task in the main experiment. For example, in the preliminary study, subjects had to judge whether *shatter* was an object (*no*) and then whether *break* was an object (*no*). In the main experiment, subjects had to judge whether *shatter* was an object (*no*) and then whether *brake* was an object (*yes*). If meaning is phonologically activated, then both meanings of a homophone should be available, and priming the incorrect meaning should increase exemplar decision latencies, errors, or both. Furthermore, as outlined above, the verification account predicts that priming should have a greater effect on exemplars with low-frequency foils, because low-frequency foils should be more likely to erroneously pass the spelling check than high-frequency foils, whereas dual-route theory predicts that priming should have a greater effect on exemplars with high-frequency foils, because high-frequency foils should be better able to compete with the exemplar than low-frequency foils. On the other hand, if meaning is activated before the influence of phonology, preceding exemplars with primes related to the foil should have no effect.

Main Experiment

Method

Subjects. Twenty-four McGill University undergraduates were paid $5 each to participate in the study. None had participated in the preliminary study. All were native speakers of English.

Stimuli. The target homophones were the 48 homophone exemplars of the categories *living thing* and *object* that were used in Experiments 1–3 (e.g., object–*brake*). The 48 words identified in the preliminary study that were semantically related to the foils served as primes. An additional 48 words, each unrelated to one of the exemplars, were chosen to serve as unrelated controls. All of the primes and unrelated controls required a *no* response. The experimental words are presented in Appendix C.

Two lists were created for each category; both lists for a category contained the 24 homophone exemplars belonging to that category. On one list they were preceded by their prime, and on the other they were preceded by their unrelated control. On each list, half of the exemplars were preceded by their prime and the other half were preceded by their unrelated control. An additional 112 filler trials were included on each list; 56 were exemplars of the category and 56 were not. They were placed randomly on the lists but not between a prime–target pair. Thus, there were two *living thing* lists and two *object* lists, each with 160 trials, 80 of which were *yes* trials and 80 of which were *no* trials. Only 7.5% of the words on a list were homophone exemplars preceded by a word semantically related to the other member of the homophone pair. Two practice lists were created, one for the *living thing* category and one for the *object* category. Half of the items on each list were members of the category and half were not.

Procedure. The subjects completed two experimental sessions, each lasting 30 min. The sessions were separated by at least 1 week. In a session, the subjects were told the name of a category and saw the practice trials followed by a list of 160 trials for that category and then were told the name of the other category and saw the practice trials and a list of 160 trials for that category. Subjects were given feedback on each practice trial to ensure that they understood the task. Subjects saw each of the four lists once. The order of presentation of the experimental lists was counterbalanced across subjects such that each list was presented in each of the four list positions six times. Words were presented on the screen one at a time and remained until the subject responded. Subjects were not told that words on some trials were semantically related to words on the subsequent trial. The intertrial interval was 1 s. Other aspects of the procedure were the same as in the previous experiments.

Results

Of the 640 trials in the experiment, only data from the 96 homophone-exemplar trials were analyzed. Before analysis, the percentage error data were arcsine transformed. A subject's response latency on a trial was only included in the analyses if the subject responded correctly to the exemplar in the primed and unprimed conditions. This criterion was met by 87.3% of responses. Another 4.5% of responses were correct but were discarded because errors were made on the exemplar in the other prime condition. A total of 1.2% of response times were greater than 1,500 ms and were replaced with times of 1,500 ms. There were three factors in the analyses; prime condition (prime vs. unrelated control), frequency on the exemplar (high vs. low), and frequency of the foil (high vs. low). These were all treated as within factors in the subject analyses. In the item analyses, prime condition was treated as a within factor, and the other two were treated as between factors.

The mean percentage errors for primed and unprimed exemplars are presented in Figure 6. There was a main effect of prime condition in the error data, $F(1, 23) = 12.55$, $p < .01$, by subjects, and $F(1, 44) = 9.01$, $p < .01$, by items. Subjects made 6.6% errors on unprimed exemplars and 9.9% errors on primed exemplars. The main effect of exemplar frequency was significant by subjects, $F(1, 23) = 11.33$, $p < .01$, but not by items, $F(1, 44) = 2.26$, $p > .05$. Subjects made 5.8% errors on high-frequency exemplars and 10.7% errors on low-frequency exemplars. The main effect of foil frequency was significant by subjects, $F(1, 23) = 39.01$, $p < .001$, but not by items, $F(1, 44) = 1.72$, $p > .05$. Subjects made 11.8% errors on exemplars with high-frequency foils and 4.8% errors on exemplars with low-frequency foils.

There was a significant interaction between prime condition and exemplar frequency by subjects, $F(1, 23) = 7.17$, $p < .05$,

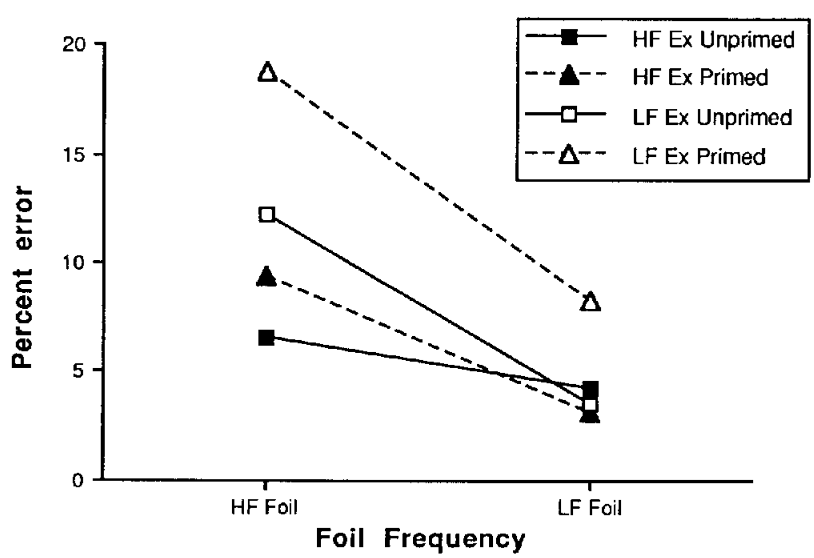

Figure 6. Mean percentage errors for primed and unprimed homophone exemplars in Experiment 5. (HF = high frequency; LF = low frequency; and Ex = exemplar.)

but not by items, $F(1, 44) = 2.08$, $p > .05$. Simple main effects tests revealed a significant effect of priming for low-frequency exemplars (5.7%), $F(1, 23) = 24.37$, $p < .001$, by subjects, and $F(1, 44) = 9.87$, $p < .01$, by items, but no effect of priming for high-frequency exemplars (0.9%; both $Fs < 1.2$). The interaction between prime condition and foil frequency approached significance by subjects, $F(1, 23) = 3.46$, $p < .08$, and was significant by items, $F(1, 44) = 4.48$, $p < .05$. Simple main effects tests revealed a significant effect of priming for exemplars with high-frequency foils (4.7%), $F(1, 23) = 12.16$, $p < .01$, by subjects, and $F(1, 44) = 13.09$, $p < .001$, by items, and an effect of priming for low-frequency foils (1.8%) that was significant by subjects, $F(1, 23) = 4.23$, $p = .05$, but not by items ($F < 1$). The triple interaction was not significant by subjects, $F(1, 23) = 1.33$, $p > .05$, or by items ($F < 1$).

In the decision latency data, there was no main effect of prime condition (both $Fs < 1$). The main effect of exemplar frequency (26 ms) was significant by subjects, $F(1, 23) = 9.64$, $p < .01$, but not by items ($F < 1$). The main effect of foil frequency (17 ms) was also significant by subjects, $F(1, 23) = 5.67$, $p < .05$, but not by items, $F(1, 44) = 1.02$, $p > .05$. None of the interactions were significant.

Discussion

The results of Experiment 5 provide little evidence for phonological activation of the meanings of high-frequency words. There was no effect of priming on either decision latencies or errors for these words even though the preliminary study demonstrated that the primes did indeed activate the foils. Had the meaning associated with the foil also been activated by the phonological representation of the exemplar, priming should have increased decision latencies, errors, or both.

It could be argued that the meanings of high-frequency words were phonologically activated, but the spelling check did not err. However, although high-frequency foils may not be likely to erroneously slip by the spelling check because their spellings are familiar, priming a low-frequency foil should have led to its being chosen for the spelling check first on some occasions, and its less familiar spelling should have resulted in more errors. But there were actually 1.1% fewer errors on high-frequency exemplars with low-frequency foils when they were primed. Also, a delay would be expected when the foil was chosen for the spelling check first, and no effect of priming was observed in the latency data. These results suggest that the meanings of high-frequency words are not phonologically activated. In contrast, there was evidence for phonological activation of the meanings of low-frequency words, because priming the foil did produce more errors on homophone exemplars.

The results of Experiment 5 further suggest that a spelling check is not performed when "trick" foil trials (e.g., object-*break*) are excluded from the experiment. Priming increased errors on exemplars with high-frequency foils but had only a small effect on exemplars with low-frequency foils. If a spelling check were being performed, the reverse pattern should have been observed. The spelling check should have been better able to avoid errors when the spelling of the nontarget member of the homophone pair was more familiar, as was found in Experiments 1–4 and in Van Orden's experiments. Subjects in Van Orden's experiments and in Experiments 1–4 probably were cautious and performed a spelling check because of the presence of homophone-foil trials in the experiments. That is, they may have had enough information to respond without performing the spelling check, but because they were aware of the presence of trick trials in which the target sounded like a member of the category, they sought additional orthographic information before responding. Thus, the results of Experiment 5 imply that the spelling check is not an obligatory step in word recognition, as Van Orden (1987; Van Orden et al., 1988, 1990) argued, but rather is a strategy used to avoid errors in a laboratory experiment.

The results of Experiment 5 support the dual-route view that a spelling check is not a normal part of the word-recognition process. For high-frequency words, meaning is activated directly form orthography, with little effect of phonology. For low-frequency words, meaning is activated from both orthography and phonology. This information is usually sufficient to prevent an error on a low-frequency homophone: The correct meaning will receive activation from both sources, but the incorrect meaning will receive activation from the phonological representation only. In the latter case, decision latencies might be slowed by the presence of a second active semantic representation, but subjects would be expected to answer correctly. However, when the foil of a homophone is primed, its activation level should exceed that of the target on some trials, particularly if the foil is a common word. Priming of the foil has no effect on high-frequency words because the meanings of these words are strongly activated from orthography before activation from their phonological representations.

Experiment 6

Experiments 1–5 suggest that there is phonological activation of the meanings of low- but not high-frequency words. One remaining objection might be that the same homophone pairs were used in all of the experiments (except for the few added in Experiment 4), and so the possibility exists that the results are specific to the categories and items used. In Experiment 6, a new category decision (Is it a verb?) and set of words were chosen to examine whether the finding of phonological effects only for low-frequency words replicates.

The design of the experiment was also different from previous experiments. In Experiments 1–4, performance on homophone foils was compared with that on spelling and homophone controls, and in Experiment 5, performance on primed exemplars was compared with performance on unprimed exemplars. In the final experiment, performance on *yes* responses to exemplars (e.g., verb–*meet*) was compared with performance on semantically similar nonhomophone words (e.g., verb–*join*). If meaning is phonologically activated, subjects should produce longer decision latencies, more errors, or both on homophones than on nonhomophone semantic controls. This design more closely approximates natural reading insofar as it does not use homophone foils and there is no priming of the other member of the homophone pair. Observing *yes* decision latencies should allow the detection of small effects of homophony if they occur. The logic of the design is the same as that of the original homophony studies conducted by Rubenstein et al. (1971) and Davelaar et al. (1978), but the use of the semantic-decision task, rather than the lexical-decision task, ensures that meaning has been consulted and avoids the problem of choosing the appropriate pseudoword distractors.

Homophone exemplars and semantic controls were matched on typicality of category membership, frequency, and length. Shoben (1982) argued that these three factors must be controlled when testing for an effect of another variable on *yes* trials in a semantic-decision experiment. In addition, homophone exemplars and semantic controls were also matched on the frequency with which they are used as verbs. This was necessary because many verbs can also be used as nouns (e.g., to *meet*, a track *meet*). It was impossible to wholly avoid these verbs; however, if homophone and control verbs are matched on their printed frequency and also on the frequency with which they are used as verbs, any differences that are observed in performance on these items cannot be attributed to differences in degree of ambiguity between the two senses. No attempt was made to have semantic controls also be spelling controls as in Experiments 1–4. Spelling controls are not necessary, as they were in those experiments, because the trials of interest are not foil trials. With foils, subjects could falsely indicate that a foil was a member of the category because of a visual similarity to the exemplar. In this experiment, however, there was an equal likelihood that subjects would falsely indicate that homophone or nonhomophone exemplars were not members because they looked like another nonmember word. Homophone exemplar targets would only be influenced by their visual similarity to the foil if the foil received activation from the phonological representation of the exemplar.

Method

Subjects. Twenty-five McGill University undergraduates were paid $3 each to participate in the study. An additional 25 subjects volunteered to fill out a typicality ratings questionnaire. All were native speakers of English.

Stimuli. There were 160 experimental words in the study. Half were homophones and half were nonhomophones (see Appendix D). The two types of words were matched as closely as possible for meaning (e.g., *meet*, *join*). All experimental words were members of the category *verb*, and the other members of the homophone pairs were not usually used as verbs (e.g., *meat*). The homophones fell into four groups: high-frequency exemplar/high-frequency foil, high-frequency exemplar/low-frequency foil, low-frequency exemplar/high-frequency foil, and low-frequency exemplar /low-frequency foil. The semantically similar controls were matched as closely as possible to the homophones for printed-word frequency, the frequency with which they appeared in the Francis and Kucera (1982) count as verbs, typicality, and length (see Table 3).

Typicality ratings for the 160 experimental words were obtained using a questionnaire based on those used by Rosch (1975). The 160 words were listed on two pages, and beside each word was a 7-point scale. Instructions on a separate page were a paraphrase of those used by Rosch, and they asked subjects to indicate for each word how good an example of the category they felt the item was by circling a number on the scale. A *1* indicated *a very good example* and a *7* indicated *a very poor example.* Rosch's instructions included an illustration of a good and a poor example of one of the categories she was interested in. The paraphrase used here was "For example, to me *jump* is a good example of a verb, whereas *believe* is a poor example of a verb. The characteristic *verb* seems more salient for *jump* than for *believe.*" Subjects were encouraged to indicate their own opinion about each item. These data were collected to ensure that the homophone exemplars and semantic matches were equally good exemplars of the category, so that any differences could be attributed to differences in homophony. None of the differences between the mean typicality ratings of homophone exemplars and semantic controls for any of the four groups were significant (all $ts < 1.5$, $ps > .15$).

Two lists were created. Half of the homophone exemplars from each group appeared on each list. The semantic controls were placed

Table 3
Frequency, Typicality, and Length Statistics for the Words Used in Experiment 6

Group	Word frequency (exemplar)	Word frequency (foil)	Verb frequency (exemplar)	Typicality	Length
HF exemplar/					
HF foil	273.4	728.1[a]	261.6	2.9	4.0
Semantic match	278.9	—	249.6	2.7	4.3
HF exemplar/					
LF foil	187.2	4.9	166.4	3.0	4.7
Semantic match	181.3	—	168.4	2.8	4.6
LF exemplar/					
HF foil	6.4	266.2	5.3	3.5	4.6
Semantic match	6.1	—	5.0	3.2	4.8
LF exemplar/					
LF foil	4.1	5.1	3.3	3.2	4.4
Semantic match	4.1	—	3.2	2.9	4.8

Note. Word frequency was calculated using the Kucera and Francis (1967) norms, and verb frequency was calculated using the norms of Francis and Kucera (1982). (See text for a description of the typicality measure.) HF = high frequency; LF = low frequency.
[a] This figure is inflated by two items; the mean frequency of the other 18 items is 331.3.

on the lists such that they did not appear on the same list as their matched homophone exemplar. An additional 240 filler trials were included. Of these, 40 were exemplars of the category *verb* and 200 were not verbs. The order of presentation of stimuli on each list was random. The lists thus had 200 trials each: half required a *yes* response and half required a *no* response. A homophone was presented on 20% of trials. Another 16 words, half of which were verbs, were chosen to serve as practice stimuli.

Procedure. The subjects completed one experimental session lasting 30 min. They were told that they were to decide whether each word presented was a verb. They were then shown the practice words and were given feedback on each trial to ensure that they understood the task. The experimental lists were presented after the practice; half of the subjects saw List 1 first and half saw List 2 first. After a short break they were shown the remaining list. Words were presented on the screen one at a time and remained until the subject responded. The intertrial interval was 1.5 s. Other aspects of the procedure were the same as in the previous experiments.

Results

Of the 400 trials in the experiment, only data from the 80 homophone exemplar and 80 semantic match *yes* trials were analyzed. Before analysis, the percentage error data were arcsine transformed. A subject's response latency on a trial was only included in the analyses if the subject responded correctly to both the homophone and its semantic control. This ensured that the same number of responses contributed to each mean. This criterion was met by 81.5% of responses. Another 8.0% of responses were correct but were discarded because the subjects made an error on the other member of the stimulus pair. A total of 3.0% of response times were greater than 1,500 ms and were replaced with times of 1,500 ms. Almost a third of these were the response times of 1 subject. There were three factors in the analyses: frequency of the exemplar (high vs. low), frequency of the foil (high vs. low), and homophony (homophone vs. semantic control). The latency and error results are shown in Figure 7.

In the decision latency data, the main effect of homophony (11 ms) was not significant, $F(1, 24) = 3.29$, $p > .05$, by subjects, and $F(1, 152) = 1.88$, $p > .05$, by items. The main effect of exemplar frequency (34 ms) was significant, $F(1, 24) = 9.78$, $p < .01$, by subjects, and $F(1, 152) = 6.83$, $p < .01$, by items. The main effect of foil frequency (4 ms) was not significant (both Fs < 1).

The interaction between homophony and exemplar frequency was significant by subjects. $F(1, 24) = 4.46$, $p < .05$, and items, $F(1, 152) = 3.77$, $p = .05$. The effect of homophony was −4 ms for high-frequency words and 26 ms for low-frequency words. The interaction between homophony and foil frequency was not significant by subjects, $F(1, 24) = 1.97$, $p > .05$, or by items, $F(1, 152) = 1.95$, $p > .05$. The triple interaction between homophony, exemplar frequency, and foil frequency was not significant (both Fs < 1).

In the error data, the main effect of homophony (2.8%) approached significance by subjects, $F(1, 24) = 3.80$, $p = .06$, but was not significant by items, $F(1, 152) = 1.58$, $p > .05$. Neither the main effect of exemplar frequency, $F(1, 24) = 1.22$, $p > .05$, by subjects ($F < 1$, by items) nor the main effect of foil frequency (both Fs < 1) was significant.

The interaction between homophony and exemplar frequency approached significance by subjects, $F(1, 24) = 3.52$, $p < .08$, but not by items ($F < 1$). The effect of homophony was 0.7% for high-frequency words and 4.8% for low-frequency words. The interaction between homophony and foil frequency was significant by subjects, $F(1, 24) = 4.54$, $p < .05$, but not by items ($F < 1$). The effect of homophony was

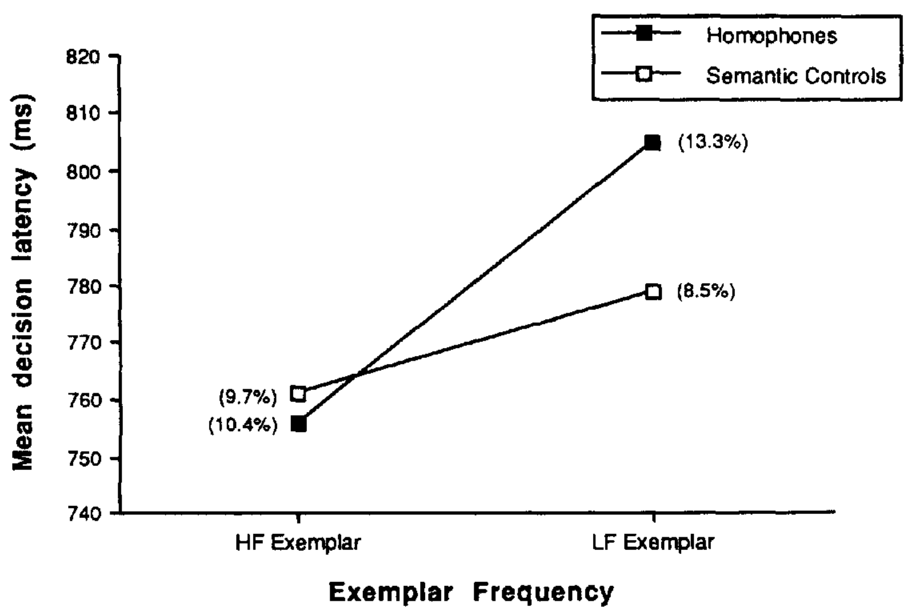

Figure 7. Mean decision latencies for homophone and semantic control exemplars (collapsed across foil frequency) in Experiment 6. (Percentage errors are in parentheses. HF = high frequency; LF = low frequency.)

1.1% when homophones had high-frequency foils and 4.4% when they had low-frequency foils. The triple interaction between homophony, exemplar frequency, and foil frequency was not significant by subjects, $F(1, 24) = 3.08$, $p > .05$, or by items ($F < 1$).

A further analysis examined the correlation between the size of the homophone effect and the orthographic similarity of the homophone exemplar and its foil. Orthographic similarity was calculated using Weber's (1970) measure. If subjects use a spelling check, and a word is more likely to falsely pass if it is orthographically similar to the exemplar, there should be a positive correlation between the size of the homophone effect in the error data and the similarity of the members of the homophone pair. When all items were included in the analysis of the error data, this relation was not observed ($r = .01$). The verification theory predicts that the correlation would occur particularly when homophone exemplars have low-frequency foils because false candidates are more likely to pass the spelling check if their spellings are unfamiliar. However, when only items with low-frequency foils were included, this relation was extremely weak ($r = .11$). And, if a low-frequency false candidate is only submitted to a spelling check if a higher frequency word has not been found to be an exemplar first, there should be a correlation between the size of the homophony effect and spelling similarity when both members of a homophone pair are low in frequency. However, this relation was again not observed when only the data from the low-frequency exemplar/low-frequency foil group were included ($r = .08$). In the latency data, these correlations were slightly higher ($r = .19$, .22, and .15 in the three analyses, respectively) but still nonsignificant.

Discussion

No evidence was found in Experiment 6 for phonological activation of the meanings of high-frequency words. On the other hand, subjects did show evidence that the meanings of low-frequency words are phonologically activated because a significant effect of homophony for low-frequency words was found in the decision latency data. The results of Experiment 6 do not provide support for the view that the meanings of words are exclusively (Van Orden, 1987; Van Orden et al., 1988) or predominantly (Van Orden et al., 1990) activated by phonological representations, nor do they provide support for the view that a spelling check is normally performed on candidates (Becker, 1976, 1980; Davelaar et al., 1978; Paap et al., 1982; Rubenstein et al., 1971; Van Orden, 1987; Van Orden et al., 1988, 1990). If such a check had been performed on phonologically activated high-frequency words, there should have been an effect of homophony in the latency data for high-frequency exemplars with high-frequency foils, because on at least some occasions the exemplar would be submitted to the spelling check only after the foil had been checked. In fact, decision latencies were 16 ms faster for homophones than semantic controls in this group. Furthermore, a strong correlation between spelling similarity and the size of the homophone effect should have been present in the data, and this was not observed.

The data instead support the time-course dual-route view that both orthographic and phonological information activate semantic representations and that the meanings of high-frequency words are highly activated by orthographic information before there is much or any activation by phonology. For low-frequency words, activation from the orthographic representation is slower and therefore weaker, and so activation from phonology does have an opportunity to contribute to the activation of the semantic representation. Although phonological activation may lead to inappropriate meanings being activated in the case of homophones, the combination of orthographic and phonological evidence for the word shown is usually enough to prevent an error. However, it may take a little longer for the correct semantic representation to dominate when other possibilities are also activated. In Experiments 1–4, there was an effect of the frequency of the other member of the homophone pair, which indicated that subjects were performing a spelling check on low-frequency words before responding. There was no effect of foil frequency in the present experiment. This suggests that the spelling check is a cautious strategy that subjects adopt when the experiment contains targets that sound like they require a *yes* response (object–*break*) but actually require a *no* response.

General Discussion

The present study addressed whether phonological information plays a role in skilled word recognition. This issue is an important one; knowing how skilled readers recognize words has broad implications for the teaching of reading skills (especially the relevance of the phonics method) and the identification of the causes of poor reading (Adams, 1990). The role of phonology in visual word recognition is perhaps the single most extensively studied issue in the reading field. All possible theoretical positions have been held at one time or another: that word recognition is necessarily phonologically mediated, necessarily direct, or both. Despite extensive study using sophisticated on-line processing methodologies, it has been difficult to achieve closure on the issue. One reason is because it has been difficult to develop methods that determine whether phonological representations are activated before or after the activation of meaning. In addition, many of the tasks that have been used are susceptible to experiment-specific strategies, making it difficult to know whether the results generalize to reading under other conditions. The present studies have overcome some of the difficulties associated with the previous studies by examining performance on homophones in a semantic-decision task. They also examined the use of subject strategies on that task.

Phonological Activation of Meaning

The results of these experiments strongly suggest that the meanings of printed words are activated by phonology only for lower frequency words. There was no evidence for phonological activation of the meanings of high-frequency words. Subjects were not more likely to make false-positive errors on high-frequency homophone foils than on spelling controls when priming from the category name was reduced. Nor was there any evidence that priming the other member of the homophone pair increased errors or decision latencies on high-frequency homophone exemplars. And finally, subjects did not produce longer decision latencies or more errors on high-frequency homophone exemplars than on nonhomophone semantic controls. The absence of effects of phonology for high-frequency words suggests that their meanings are activated directly on the basis of visual information.

In contrast, phonological effects were observed for lower frequency words in each experiment. There were more false-positive errors on low-frequency homophone foils than on spelling controls; priming the other member of the homophone pair caused an increase in errors on low-frequency homophone exemplars, and decision latencies for low-frequency homophone exemplars were longer than for nonhomophonic semantic controls. Evidence that these phonological effects arise from processing before the activation of meaning comes from the observation in Experiment 2 of similar false-positive rates for low-frequency homophone and pseudohomophone foils. Furthermore, there was evidence that the use of a phonological code in reading these words is not an optional strategy for skilled readers. In Experiment 3, more errors were made on low-frequency homophone foils than on spelling controls even when subjects were discouraged from using a phonological code by including a large proportion of homophones.

Using the lexical-decision task, Andrews (1982), Davelaar et al. (1978), Waters and Seidenberg (1985), and Waters, Seidenberg, and Bruck (1984) also found evidence for phonological recoding only for low-frequency words. However, by varying the properties of the word and nonword stimuli, it is also possible to eliminate these effects (Davelaar et al., 1978; Seidenberg et al., 1984; Waters & Seidenberg, 1985). The conclusion that meaning is phonologically activated for low-frequency but not high-frequency words can be made more strongly on the basis of the present experiments. One reason is that the task used here—semantic-decision—requires the subject to focus on the meanings of words, whereas the lexical-decision task does not. The second reason is that evidence for phonological activation of meaning was observed only for low-frequency words in Experiment 6, which did not include either pseudoword or foil trials. Thus, the results are more readily generalizable to normal reading than are the results from lexical-decision experiments that are susceptible to special strategies because nonwords must be included. Finally, the present experiments used a homophony manipulation instead of the spelling–sound consistency manipulation used in some previous studies of the role of phonology (Andrews, 1982; Bauer & Stanovich, 1980; Coltheart, Besner, Jonasson, & Davelaar, 1979; Parkin, 1982; Seidenberg et al., 1984; Stanovich & Bauer, 1978; Waters & Seidenberg, 1985; Waters et al., 1984). Recent work suggests that a failure to find a difference between inconsistent and consistent high-frequency words does not necessarily indicate that the meanings of these words are activated directly on a visual basis. The spelling–sound translation system proposed by Seidenberg and McClelland (1989) can derive the correct pronunciations of exception words, unlike grapheme–phoneme correspondence rules (Coltheart, 1978), and after sufficient training, the phonological representations of high-frequency exception words are computed as easily as those of regular words. A difference should, however, have been observed between high-frequency homophones and nonhomophones if meanings are activated by phonology because homophones will activate more than one semantic representation.

The Spelling-Check Procedure

The spelling check plays a central role in the word recognition theory proposed by Van Orden and colleagues (Van Orden, 1987; Van Orden et al., 1988). A mechanism of this kind is needed to reconcile the fact that people identify homophones correctly with their claim that meaning is exclusively or predominantly activated by phonological representations. The results of the present experiments suggest that a spelling check is not usually performed, but rather is a strategy that is used when the experiment contains foil trials, as in Experiments 1–4. In Experiments 1–3, subjects made fewer errors on low-frequency homophone foils when they had high-frequency exemplars than when they had low-frequency exemplars. This is consistent with the view that a spelling check was performed, because subjects were better able to avoid errors when the spelling of the exemplar was familiar. In Experiment 4, subjects made more errors on low-frequency homophone foils with similarly spelled exemplars than on foils with dissimilarly spelled exemplars. This is consistent with the hypothesis that exemplars are more likely to falsely slip by the spelling check if they are spelled like the target foil. These studies support the view that a spelling check is performed in experiments containing foil trials. In Experiment 5, however, no foil trials were included, and in that experiment, subjects made more errors on homophone exemplars when the other member of the homophone pair was primed and was high in frequency than when the other member of the pair was primed and was low-frequency. If a spelling check had been performed, subjects should have been better able to avoid errors when the spelling of the other member of the homophone pair was familiar because it is high in frequency. Furthermore, in Experiment 6, the size of the homophony effect in the error data was not correlated with the similarity between the exemplar and the other member of the homophone pair. If a spelling check was being performed, subjects should have been more likely to make errors on homophone exemplars when the two were similarly spelled.

The failure to find evidence for the use of a spelling check when no foil trials were included suggests that activation from orthography to meaning is normally strong enough to allow readers to avoid making errors on low-frequency homophones, although they will be slowed by the competition from the other member of the homophone pair. When foil trials are included, subjects are more cautious and perform a spelling check to avoid making false-positive errors. These conclusions are consistent with previous research. The Van Orden studies (Van Orden, 1987; Van Orden et al., 1988) all involved homophone foils, and the results were explained in terms of a spelling-check mechanism. Van Orden (1987) noted that other evidence for this mechanism comes from studies that included pseudohomophone foils among the stimuli (Becker, 1976, 1980; Becker & Killion, 1977; Schvaneveldt and McDonald, 1981). Another study providing evidence for a spelling-check mechanism (Norris, 1984) included pairs of trials, such as *bread–batter*, that also may induce a cautious strategy.

Implications for Theories of Word Recognition

The results of Experiments 1–6 do not support Van Orden's (1987; Van Orden et al., 1988) hypothesis or the hypothesis of Rubenstein et al. (1971) that bottom-up activation of candidates occurs exclusively by way of their phonological representations and that a spelling check is performed on the most highly activated candidates. They instead support the dual-route view that both orthographic and phonological pathways to meaning exist. Direct activation of meaning from orthography was observed to occur for high-frequency words, and along with activation from phonology, it contributed to the activation of the meanings of low-frequency words, obviating the need to routinely engage in a spelling check to disambiguate homophones.

Although the results of the present experiments are consistent with the broad dual-route approach, they are not consistent with some proposals made by dual-route theorists. Coltheart (1978) argued that in skilled reading, processing along the visual route was nearly always completed before processing along the phonological route. The results of the present experiments suggest that this is true only for high-frequency words. At the other extreme, Van Orden et al. (1990) claimed that the phonological route will predominate over the direct route in the computation of a semantic code. The results of the present experiments suggest that the phonological route only contributes to the activation of the meanings of low-frequency words. A common assumption of dual-route models (e.g., Coltheart, 1978; Davelaar et al., 1978; Hawkins et al., 1976; McQuade, 1981) is that use of the phonological route is a strategy under control of the subject. However, results of the present experiments suggest instead that it is the spelling check that is strategic. These conclusions are consistent with the finding that phonological information appears to be activated quickly and automatically (Bakan & Alperson, 1967; Dalrymple-Alford, 1972; Humphreys et al., 1982; Perfetti et al., 1988; Tanenhaus et al., 1980).

The fact that the extent to which there is activation of meaning from phonology depends on word frequency can be explained in terms of the time course of processing along the direct and phonological routes. We will use the Seidenberg and McClelland (1989) model to develop this idea (Figure 8), although it has been proposed in somewhat different form by others (e.g., Meyer et al., 1974; Paap et al., 1987; Seidenberg, 1985a, 1985b). As in other dual-route models, there are two pathways to meaning: one directly from orthography and the other from orthography to phonology to meaning. The computation from orthography to phonology has been implemented as a working simulation model. The model differs from previous accounts in that it does not include representations of individual words; rather, the spelling, pronunciation, and meaning of a word are represented by patterns of

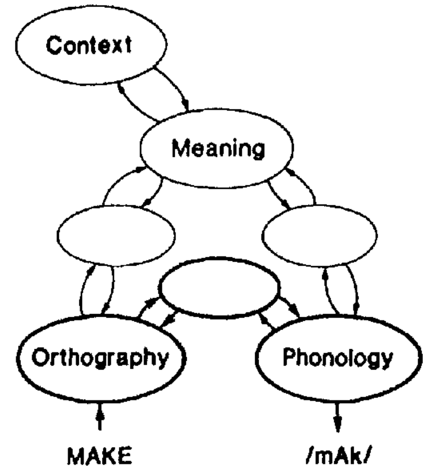

Figure 8. The Seidenberg and McClelland (1989) model. (Reprinted from "A Distributed, Developmental Model of Word Recognition and Naming" by M. S. Seidenberg and J. L. McClelland, 1989, *Psychological Review, 96,* p. 526. Copyright 1989 by the American Psychological Association.)

activation distributed across simple processing units encoding each of these types of information. The weights on connections between orthographic and phonological units encode facts about the correspondences between spelling and sound. Theoretically, there are also weighted connections between orthographic and semantic units and between phonological and semantic units that encode the correspondences between these representations, although these parts of the model have not been implemented. In the simulation, the weights are initialized with small random values and then are modified by means of the back-propagation learning algorithm (Rumelhart, Hinton, & Williams, 1986) during a training phase in which the model is exposed to a large number of words and their pronunciations. The weights gradually come to encode facts about the consistency of spelling–sound correspondences in the training corpus, so that the correct phonological representation of a word is computed when an orthographic pattern is presented. Van Orden et al. (1990) proposed a similar view.

In the Seidenberg and McClelland (1989) model, the presentation of a visual stimulus is assumed to initiate the spread of activation outward from the orthographic units. Activation spreads directly to the semantic units and directly to the phonological units; there is secondary spread of activation from the phonological units to the semantic units. The time course of processing is determined by the values of the weights on the different sets of connections. The semantic representation of a word thus builds up over time as activation spreads from the two sources. This represents an important contrast to other dual-route accounts in which meaning is assumed to be *accessed* on the basis of one or the other of the two autonomous processing routines. In horse-race models, for example (Forster & Chambers, 1973; Meyer et al., 1974; Paap et al., 1987), the direct and phonologically mediated pathways operate in parallel but autonomously; the race between them determines which provides access to meaning. The Seidenberg and McClelland model incorporates the idea of two parallel ways of activating meaning; however, the different processes can jointly contribute to the activation of distributed semantic representations. Within this model, computing a representation over the semantic units corresponds to the process of accessing the meaning of a word in earlier models. The secondary spread of activation from phonological to semantic nodes corresponds to phonological mediation. It is not entirely appropriate to apply the term *phonological mediation* to this process, however, because the term carries further theoretical implications that are being called into question. In earlier models, the orthographic stimulus was recoded into its full and explicit phonological representation, which was then used to search lexical memory (Rubenstein et al., 1971). In the Seidenberg and McClelland model, there can be partial activation of distributed representations; thus, orthographic input can partially activate phonological representations, and they in turn can partially activate semantic representations. Hence, activation of meaning by phonology does not require computing the entire phonological representation of a word. For this reason, we prefer to term this process the *phonological activation of meaning.*

Whether there is spread of activation from phonology to meaning depends on the time course of processing along the two pathways. In the case of higher frequency words, our empirical results suggest that meaning is principally activated directly from orthography. This means that the semantic nodes have become sufficiently activated to support performance on the categorization task before activation has spread from orthography to phonology to meaning. For lower frequency words, the time course of processing is such that there is more activation of meaning from phonology. Apparently, the semantic nodes do not become sufficiently activated on the basis of orthography alone to permit a response, which must await the arrival of secondary activation from phonology. Thus, this model suggests that the extent to which phonology influences the activation of meaning can differ in degree depending on the timing of this process relative to the activation of meaning directly from orthography.

Within the framework illustrated in Figure 8, the key question concerns the time course of processing, that is, the rate at which different types of output units are activated. Several computations—from orthography to phonology, from orthography to semantics, and from phonology to meaning—are assumed to occur in parallel. The characteristics of these computations are determined by the weights on connections between units. Seidenberg and McClelland's (1989) simulation model is quite limited in that only the computation from orthography to phonology is implemented; there is no representation of meaning. Moreover, the implemented model is not a real-time system; orthographic and phonological output are computed in a single step, rather than in the cascaded manner we assume is true in reality. Hence, little can be said about detailed aspects of the time course of processing, specifically, how much activation will spread from phonology to semantics for different words. This is an important issue that needs to be addressed in future simulation models. Other factors being equal, however, it is clear that there should be more secondary activation of meaning from phonology when the computation from orthography to meaning is itself very slow. This might occur, for example, when the input word is low in frequency. Some of these words may be more familiar from speech than from reading; hence, meaning can be activated only through phonology. Perhaps the limiting case is provided by pseudohomophones, such as *sute*, for which the computation from orthography to meaning will produce very little semantic output because the string is a nonword; however, such nonwords could produce activation of meaning through phonology. Similar effects may obtain if the reader's decoding skills are poor. By hypothesis, the poor reader computes meanings from orthography very slowly, allowing more time for activation from the indirect orthography–phonology–meaning route. Although this account needs to be tested in a simulation model that includes both routes, it represents the beginning of a computational account of why phonological activation of meaning is apparently associated with lower frequency words.

The theory of Van Orden et al. (1990) differs most from Seidenberg and McClelland's (1989) model in its assumptions about the relative influence of the two processing routes on the activation of semantic representations. Van Orden et al. (1990) claimed that the nature of the computation between orthography and phonology will result in this route having a predominant influence in the computation of all semantic codes. The results for high-frequency words in the experiments presented here suggest that this view is not correct.

In summary, the Seidenberg and McClelland (1989) model incorporates some aspects of previous dual-route and horse-race models, but differs from them in critical respects. Previously, it was assumed that the meaning of a word could only be accessed by first locating an entry in an orthographic or phonological lexicon (or activating a logogen). Processing involved a parallel race between autonomous recognition processes. In the Seidenberg and McClelland model, determining the meaning of a word does not require accessing a stored orthographic or phonological code; rather, the codes for words are *computed* each time they are read. The same sets of units are used to represent all words; all that differs is the pattern of activation produced by each word. The computed semantic code develops over time as activation accrues from different sources.

According to this theory, then, computing the code of a word—its meaning or pronunciation, for example—is a constraint satisfaction problem (see McClelland, Rumelhart, & Hinton, 1986). The code that is computed satisfies a large number of simultaneous constraints that, in the model, are represented by the weights on connections between units. In the case of meaning, the extent to which the computed code is determined by orthographic or phonological knowledge depends on the settings of the weights. Although we have focused on these two sources of activation, there may be additional activation from other sources, such as the linguistic and nonlinguistic contexts in which a word occurs. This is clearly seen in the case of homonyms, such as *tire*, in which multiple meanings are associated with a single spelling and single pronunciation, and disambiguation requires input from contextual sources (see Cottrell, 1988; Kawamoto, 1988;

Swinney, 1979; Tanenhaus, Leiman, & Seidenberg, 1979, for accounts of this process). Similarly, determining the correct pronunciation of a homograph, such as *wind* or *bass*, requires activation from semantics to phonology, and the semantic representation itself may be determined in part by input from the context. In summary, various sources of information together determine the meaning of a word (Barsalou, 1987; Schwanenflugel & Shoben, 1985); the contributions from different sources depend on properties of the word and the context in which it occurs.

Conclusions

The results of the present experiments suggest that phonological information does play a role in the activation of meanings of words in skilled silent reading. This role is somewhat greater than assumed by researchers such as Coltheart (1978) and somewhat less than assumed by researchers such as Van Orden (1987; Van Orden et al., 1988, 1990). One reason for the uncertainty regarding the role of phonological information is that the tasks that have been used to explore this issue are vulnerable to task-specific strategies. Lexical-decision studies sometimes underestimate the effect of phonology because the task does not always require the activation of meaning, whereas semantic-decision studies sometimes overestimate its effect because of priming from the category name. When such task-specific strategies are factored out, it appears that phonological activation of meaning occurs only for lower frequency words. Phonology is therefore relevant to the large number of words that occur with relatively low frequency, but not to the smaller number of high-frequency words that account for most of the tokens in actual texts. Where phonology appears to play a much greater role is in the acquisition of word-recognition skills (Adams, 1990). Beginning readers depend much more heavily on phonological activation of meaning; taken with our results, this suggests that the transition to skilled reading is marked by greater reliance on direct activation of meaning. Being able to use phonological information may facilitate the acquisition of word-recognition skills; in skilled reading, it may continue to facilitate the recognition of infrequent words. These outcomes are possible because of the invention of writing systems that encode phonological information.

References

Adams, M. J. (1990). *Beginning to read: Thinking and learning about print.* Cambridge, MA: MIT Press.

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

Bakan, P., & Alperson, B. (1967) Pronounceability, attensity, and interference in the color-word test. *American Journal of Psychology, 80,* 416–420.

Balota, D. A., & Chumbley, J. I. (1984). Are lexical decisions a good measure of lexical access? The role of word-frequency in the neglected decision stage. *Journal of Experimental Psychology: Human Perception and Performance, 10,* 340–357.

Banks, W. P., Oka, E., & Shugarman, S. (1981). Recoding of printed words to internal speech: Does recoding come before lexical access? In O. J. L. Tzeng & H. Singer (Eds.), *Perception of print* (pp. 137–170). Hillsdale, NJ: Erlbaum.

Baron, J. (1985). Back to basics. *The Behavioral and Brain Sciences, 8,* 706.

Barsalou, L. W. (1987). The instability of graded structure: Implications for the nature of concepts. In U. Neisser (Ed.), *Concepts and conceptual development* (pp. 101–140). Cambridge, England: Cambridge University Press.

Bauer, D. W., & Stanovich, K. E. (1980). Lexical access and the spelling-to-sound regularity effect. *Memory and 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.

Becker, C. A., & Killion, T. H. (1977). Interaction of visual and cognitive effects in word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 3,* 389–401.

Besner, D., Davelaar, E., Alcott, D., & Parry, P. (1984). Wholistic reading of alphabetic print: Evidence from the FDM and the FBI. In L. Henderson (Ed.), *Perspectives from cognitive psychology, neuropsychology, and linguistics* (pp. 121–135). Hillsdale, NJ: Erlbaum.

Brown, G. D. A. (1987). Resolving inconsistency: A computational model of word naming. *Journal of Memory and Language, 26,* 1–23.

Carr, T. H., & Pollatsek, A. (1985). Recognizing printed words: A look at current models. In D. Besner, T. G. Waller, & E. M. MacKinnon (Eds.), *Reading research: Advances in theory and practice* (Vol. 5, pp. 1–82). San Diego, CA: Academic Press.

Clark, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior, 12,* 335–359.

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

Coltheart, M., (1980). Reading, phonological recoding, and deep dyslexia. In M. Coltheart, K. Patterson, & J. C. Marshall (Eds.), *Deep dyslexia* (pp. 197–226). London: Routledge & Kegan Paul.

Coltheart, M., Besner, D., Jonasson, J. T., & Davelaar, E. (1979). Phonological encoding in the lexical decision task. *Quarterly Journal of Experimental Psychology, 31,* 489–507.

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). Hillsdale, NJ: Erlbaum.

Cottrell, G. W. (1988). A model of lexical access of ambiguous words. In S. Small, G. Cottrell, & M. Tanenhaus (Eds.), *Lexical ambiguity resolution: Perspectives from psycholinguistics, neuropsychology, and artificial intelligence* (pp. 179–194). San Mateo, CA: Kaufmann.

Dalrymple-Alford, E. C. (1972). Sound similarity and color-word interference in the Stroop task. *Psychonomic Science, 28,* 209–210.

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

Fischler, I. (1977). Semantic facilitation without association in a lexical decision task. *Memory and Cognition, 5,* 335–339.

Forster, K. I. (1988, April). *On knowing how many entries.* Paper presented at the Conference on Semantic Ambiguity, Adelphi University, Garden City, NY.

Forster, K. I., & Chambers, S. M. (1973). Lexical access and naming time. *Journal of Verbal Learning and Verbal Behavior, 12,* 627–635.

Francis, W. N., & Kucera, H. (1982). *Frequency analysis of English* *usage: Lexicon and grammar.* Boston: Houghton-Mifflin.

Gordon, B. (1983). Lexical access and lexical decision: Mechanisms of frequency sensitivity. *Journal of Verbal Learning and Verbal Behavior, 22,* 24–44.

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

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.* San Diego, CA: Academic Press.

Humphreys, G. W., Evett, L. J., & Taylor, D. E. T. (1982). Automatic phonological priming in visual word recognition. *Memory and Cognition, 10,* 576–590.

Hung, D. L., & Tzeng, O. J. L. (1981). Orthographic variations and visual information processing. *Psychological Bulletin, 90,* 377–414.

Johnson, N. F. (1975). On the function of letters in word identification: Some data and a preliminary model. *Journal of Verbal Learning and Verbal Behavior, 14,* 17–29.

Kawamoto, A. H. (1988). Distributed representations of ambiguous words and their resolution in a connectionist network. In S. Small, G. Cottrell, & M. Tanenhaus (Eds.), *Lexical ambiguity resolution: Perspectives from psycholinguistics, neuropsychology, and artificial intelligence* (pp. 195–228). San Mateo, CA: Kaufmann.

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

Laberge, D., & Samuels, J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology, 6,* 293–323.

McClelland, J. L., Rumelhart, D. E., & Hinton, G. E. (1986). The appeal of parallel distributed processing. In D. E. Rumelhart & J. L. McClelland (Eds.), *Parallel distributed processing: Explorations in the mictrostructures of cognition* (Vol. 1, pp. 3–44). Cambridge, MA: MIT Press.

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. (1975, November). *Orthographic vs. phonemic processing of printed words.* Paper presented at the meeting of the Psychonomic Society, Denver, CO.

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

Meyer, D. E., Schvaneveldt, R. W., & Ruddy, M. G. (1974). Functions of graphemic and phonemic codes in visual word recognition. *Memory and Cognition, 2,* 309–321.

Monsell, S., Doyle, M. C., & Haggard, P. N. (1989). Effects of frequency on visual word recognition tasks: Where are they? *Journal of Experimental Psychology: General, 118,* 43–71.

Morton, J. (1969). Interaction of information in word recognition. *Psychological Review, 76,* 165–178.

Neely, J. H. (1991). Semantic priming effects in visual word recognition: A selective review of current findings and theories. In D. Besner & G. Humphreys (Eds.), *Basic processes in reading: Visual word recognition* (pp. 264–336). Hillsdale, NJ: Erlbaum.

Neely, J. H., Keefe, D. E., & Ross, K. L. (1989). Semantic priming in the lexical decision task: Roles of prospective prime-generated expectancies and retrospective semantic matching. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 1003–1019.

Norris, D. (1984). The mispriming effect: Evidence of an orthographic check in the lexical decision task. *Memory and Cognition, 5,* 470–476.

Paap, K. R., McDonald, J. E., Schvaneveldt, R. W., & Noel, R. W. (1987). Frequency and pronounceability in visually presented naming and lexical decision tasks. In M. Coltheart (Ed.), *Attention and performance XII: The psychology of reading* (pp. 221–243). Hillsdale, NJ: Erlbaum.

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

Parkin, A. J. (1982). Phonological recoding in lexical decision: Effects of spelling-to-sound regularity depend on how regularity is defined. *Memory and Cognition, 10,* 43–53.

Patterson, K. E., & Morton, J. (1985). From orthography to phonology: An attempt at an old interpretation. In K. E. Patterson, J. C. Marshall, & M. Coltheart (Eds.), *Surface dyslexia: Neuropsychological and cognitive studies of phonological reading* (pp. 335–359). Hillsdale, NJ: Erlbaum.

Perfetti, C. A., Bell, L. C., & Delaney, S. M. (1988). Automatic (prelexical) phonetic activation in silent word reading: Evidence from backward masking. *Journal of Memory and Language, 27,* 59–70.

Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General, 104,* 192–233.

Rozin, P., & Gleitman, L. R. (1977). The structure and acquisition of the alphabetic principle. In A. S. Reber & D. L. Scarborough (Eds.), *Toward a psychology of reading* (pp. 55–142). Hillsdale, NJ: Erlbaum.

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., Hinton, G. E., & Williams, R. J. (1986). Learning internal representations by error propagation. In D. E. Rumelhart & J. L. McClelland (Eds.), *Parallel distributed processing: Explorations in the mictrostructures of cognition* (Vol. 1, pp. 318–362). Cambridge, MA: MIT Press.

Rumelhart, D. E., & Siple, P. (1974). Process of recognizing tachistoscopically presented words. *Psychological Review, 81,* 99–118.

Schvaneveldt, R. W., & McDonald, J. E. (1981). Semantic context and the encoding of words: Evidence for two modes of stimulus analysis. *Journal of Experimental Psychology: Human Perception and Performance, 7,* 673–687.

Schwanenflugel, P. J., & Shoben, E. J. (1985). The influence of sentence constraint on the scope of facilitation for upcoming words. *Journal of Memory and Language, 24,* 232–252.

Seidenberg, M. S. (1985a). The time course of information activation and utilization in visual word recognition. In D. Besner, T. G. Waller, & E. M. MacKinnon (Eds.), *Reading research: Advances in theory and practice* (Vol. 5, pp. 199–252). San Diego, CA: Academic Press.

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

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review, 96,* 523–568.

Seidenberg, M. S., Waters, G. S., Barnes, M. A., & Tanenhaus, M. K. (1984). When does irregular spelling or pronunciation influence word recognition? *Journal of Verbal Learning and Verbal Behavior, 23,* 383–404.

Seidenberg, M. S., Waters, G. S., Sanders, M., & Langer, P. (1984). Pre- and postlexical loci of contextual effects on word recognition. *Memory and Cognition, 12,* 315–328.

Shallice, T., Warrington, E. K., & McCarthy, R. (1983). Reading without semantics. *Quarterly Journal of Experimental Psychology,*

35A, 111–138.
Shoben, E. J. (1982). Semantic and lexical decisions. In C. R. Puff (Ed.), *Handbook of research methods in human memory and cognition* (pp. 287–314). San Diego, CA: Academic Press.
Smith, F. (1971). *Understanding reading: A psycholinguistic analysis of reading and learning to read.* New York: Holt, Rinehart & Winston.
Spoehr, K. T., & Smith, E. 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 and Cognition, 6,* 410–415.
Swinney, D. A. (1979). Lexical access during sentence comprehension: (Re)consideration of context effects. *Journal of Verbal Learning and Verbal Behavior, 18,* 645–659.
Tanenhaus, M. K., Flanigan, H., & Seidenberg, M. S. (1980). Orthographic and phonological activation in auditory and visual word recognition. *Memory and Cognition, 8,* 513–520.
Tanenhaus, M. K., Leiman, J. L., & Seidenberg, M. S. (1979). Evidence for multiple stages in the processing of ambiguous words in syntactic contexts. *Journal of Verbal Learning and Verbal Behavior, 18,* 427–440.
Tweedy, J. R., Lapinski, R. H., & Schvaneveldt, R. W. (1977). Semantic-context effects on word recognition: Influence of varying the proportion of items presented in appropriate context. *Memory and Cognition, 5,* 84–89.
Underwood, G., & Thwaites, S. (1982). Automatic phonological coding of unattended printed words. *Memory and Cognition, 10,* 434–442.
Van Orden, G. C. (1987). A ROWS is a ROSE: Spelling, sound, and reading. *Memory and Cognition, 15,* 181–198.
Van Orden, G. C., Johnston, J. C., & Hale, B. L. (1988). Word identification in reading proceeds from spelling to sound to meaning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 371–386.
Van Orden, G. C., Pennington, B. F., & Stone, G. O. (1990). Word identification in reading and the promise of a subsymbolic psycholinguistics. *Psychological Review, 97,* 488–522.
Waters, G. S., & Seidenberg, M. S. (1985). Spelling-sound effects in reading: Time course and decision criteria. *Memory and Cognition, 13,* 557–572.
Waters, G. S., Seidenberg, M. S., & Bruck, M. (1984). Children's and adults' use of spelling-sound information in three reading tasks. *Memory and Cognition, 12,* 293–305.
Weber, R. M. (1970). A linguistic analysis of first-grade reading errors. *Reading Research Quarterly, 5,* 427–451.

Appendix A

Stimuli Used in Experiments 1–3

Exemplar	Foil	Spelling control	Exp. 1 category	Exp. 2 category
		High-frequency foil/high-frequency exemplar		
principal	principle	municipal	school employee	living thing
son	sun	sin	male relative	living thing
bear	bare	beat	hibernating animal	living thing
male	mail	mile	biological category	living thing
patients	patience	patent	people found in hospitals	living thing
residents	residence	resistance	citizens	living thing
cent	sent	count	type of money	object
clothes	close	claims	stored in closets	object
meat	meet	mean	dinner food	object
plane	plain	play	air vehicle	object
presents	presence	preserve	typically at a birthday party	object
road	rode	round	used by travellers	object
		High-frequency foil/low-frequency exemplar		
beech	beach	beer	type of tree	living thing
knight	night	knife	distinguished man	living thing
nun	none	non	religious person	living thing
seller	cellar	secret	store personnel	living thing
prophet	profit	project	Biblical person	living thing
boarder	border	broader	paying resident	living thing
sail	sale	soul	part of a boat	object
ladder	latter	labor	painter's equipment	object
pail	pale	pain	cleaning equipment	object
axe	acts	age	woodsman's tool	object
throne	thrown	throat	monarch's object	object
brake	break	brave	car part	object

Appendix A (*Continued*)

Exemplar	Foil	Spelling control	Exp. 1 category	Exp. 2 category
		Low-frequency foil/high-frequency exemplar		
guys	guise	guts	group of men	living thing
horse	hoarse	hose	farm animal	living thing
prince	prints	prance	royalty	living thing
rose	rows	robe	garden flower	living thing
guest	guessed	gust	people in hotels	living thing
boy	buoy	bog	young person	living thing
board	bored	boast	construction material	object
bread	bred	brood	baked item	object
gate	gait	gale	type of barrier	object
ring	wring	rinse	type of jewelry	object
ball	bawl	bail	child's toy	object
sign	sine	sigh	information medium	object
		Low-frequency foil/low-frequency exemplar		
flea	flee	flex	parasite	living thing
fowl	foul	foil	edible birds	living thing
toad	towed	trod	amphibian	living thing
whale	wail	whack	sea mammal	living thing
baron	barren	bargain	nobleman	living thing
pigeon	pidgin	piston	city dwelling bird	living thing
sword	soared	sworn	weapon	object
pole	poll	poke	tent part	object
medal	meddle	medley	type of award	object
urn	earn	urea	container	object
lens	lends	lent	optical device	object
bridle	bridal	brittle	horseback riding equipment	object
		Pseudohomophone foil/high-frequency exemplar		
writer	riter	writed	person in book industry	living thing
trees	treeze	treem	type of vegetation	living thing
daughter	dotter	daupher	female relative	living thing
soldiers	soljers	soltier	military personnel	living thing
dog	dawg	dag	domestic animal	living thing
chief	cheef	chiel	type of leader	living thing
boat	bote	boam	water vehicle	object
table	tabel	tadle	type of furniture	object
rock	rawk	roch	natural earth formation	object
phone	phoan	phand	means of communication	object
shoes	shews	shoss	footwear	object
scale	scail	scalm	measuring device	object
		Pseudohomophone foil/low-frequency exemplar		
eagle	eagel	eaple	predatory bird	living thing
fox	focks	fow	canine	living thing
poppy	paupy	poggy	symbolic flower	living thing
worm	wirm	wurn	legless animal	living thing
beetle	beatel	beelet	crawling insect	living thing
cod	caud	col	fish	living thing
purse	perse	porse	fashion accessory	object
drawer	droar	drawen	part of furniture	object
spear	speer	spean	hunting equipment	object
nail	nale	noil	type of fastener	object
skate	scate	skote	hockey equipment	object
stove	stoave	stoze	household appliance	object

Note. The homophone controls are not shown; they are simply the foil words paired with one of the other categories. Exp. = experiment.

(*Appendixes B and C follow on next page*)

Appendix B

Words Used in Experiment 4

Exemplar	Foil	Spelling control	Category
High similarity between exemplar and foil			
heroine	heroin	heroic	living thing
baron	barren	bargain	living thing
fowl	foul	foil	living thing
serf	surf	scarf	living thing
pigeon	pidgin	piston	living thing
toad	towed	toyed	living thing
boulder	bolder	balder	object
bridle	bridal	brittle	object
cord	chord	cod	object
lens	lends	leans	object
altar	alter	ajar	object
pearl	purl	peril	object
Low similarity between exemplar and foil			
lynx	links	lanky	living thing
fairy	ferry	fancy	living thing
hawk	hock	haul	living thing
idol	idle	idiom	living thing
doe	dough	dot	living thing
whale	wail	warn	living thing
rack	wrack	wreck	object
medal	meddle	medley	object
sword	soared	seared	object
mast	massed	marred	object
urn	earn	urea	object
sleigh	slay	slam	object

Appendix C

Words Used in Experiment 5

Homophone exemplar	Related prime	Unrelated prime	Category
High-frequency exemplar/high-frequency foil			
principal	ethic	bustle	living thing
son	star	magic	living thing
bear	exposed	launch	living thing
male	letters	echo	living thing
patients	serenity	phase	living thing
residents	dwelling	option	living thing
cent	shipped	bash	object
clothes	fasten	regret	object
meat	join	gamble	object
plane	obvious	aggravate	object
presents	existence	nimble	object
road	sat	health	object

Appendix C (*Continued*)

Homophone exemplar	Related prime	Unrelated prime	Category
	High-frequency exemplar/low-frequency foil		
guys	semblance	parallel	living thing
horse	raspy	reason	living thing
prince	photographs	hoot	living thing
rose	columns	malice	living thing
guest	estimated	pardon	living thing
boy	beacon	liberty	living thing
board	restless	past	object
bread	born	jaunt	object
gate	gallop	bite	object
ring	twist	baffle	object
ball	cry	hint	object
sign	angle	antic	object
	Low-frequency exemplar/high-frequency foil		
beech	shore	babble	living thing
knight	evening	pinch	living thing
nun	few	league	living thing
seller	basement	riddle	living thing
prophet	revenue	moist	living thing
boarder	outline	myth	living thing
sail	discount	jolt	object
ladder	following	passion	object
pail	colorless	bungle	object
axe	pretends	eclipse	object
throne	tossed	focus	object
brake	shatter	bold	object
	Low-frequency-exemplar/low-frequency foil		
flea	run	advantage	living thing
fowl	polluted	glory	living thing
toad	pulled	anxious	living thing
whale	weep	density	living thing
baron	arid	accident	living thing
pigeon	slang	tune	living thing
sword	glided	blame	object
pole	opinion	lucky	object
medal	intrude	brief	object
urn	acquire	blink	object
lens	donates	bellow	object
bridle	wedding	amount	object

(*Appendix D follows on next page*)

Appendix D

Words Used in Experiment 6

High-frequency foil		Low-frequency foil	
Homophone exemplar	Semantic control	Homophone exemplar	Semantic control
High-frequency words			
buy	shop	been	has
hear	speak	walk	run
knew	said	caught	shook
made	used	seem	appear
won	lost	rain	snow
led	fled	taught	fought
write	draw	sign	agree
roll	push	stayed	watched
see	look	shoot	kill
threw	struck	flew	slid
wait	seek	raise	lift
wore	hung	flow	fill
passed	reached	loan	grant
do	go	tied	fixed
seen	felt	shear	split
meet	join	pause	rush
rode	drove	cast	pitch
close	open	beat	hit
sent	kept	break	burst
sell	trade	thrown	swung
Low-frequency words			
blew	stung	alter	amend
counsel	advise	creak	squeak
haul	lug	maul	flog
hire	quit	peer	squint
owe	lend	savor	crave
pour	stir	leak	drip
sew	rip	tow	drag
steal	raid	purl	knit
weigh	ponder	pray	preach
whine	plead	tease	taunt
err	botch	slay	stab
stares	glares	tacked	nailed
mourning	grieving	retch	gasp
lessen	loosen	bury	dig
pare	chop	heal	mend
cite	quote	earn	reap
bawl	yell	meddle	intrude
bred	sired	soar	glide
bored	amused	wail	moan
daze	stun	flee	sprint

Received January 23, 1991
Revision received June 11, 1991
Accepted June 14, 1991 ■