

Rhyme decisions to spoken words and nonwords

JAMES M. McQUEEN

MRC Applied Psychology Unit, Cambridge, England

Lexical effects in auditory rhyme-decision performance were examined in three experiments. Experiment 1 showed reliable lexical involvement: rhyme-monitoring responses to words were faster than rhyme-monitoring responses to nonwords; and decisions were faster in response to high-frequency as opposed to low-frequency words. Experiments 2 and 3 tested for lexical influences in the rejection of three types of nonrhyming item: words, nonwords with rhyming lexical neighbors (e.g., *jop* after the cue *rob*), and nonwords with no rhyming lexical neighbor (e.g., *vop* after *rob*). Words were rejected more rapidly than nonwords, and there were reliable differences in the speed and accuracy of rejection of the two types of nonword. The advantage for words over nonwords was replicated for positive rhyme decisions. However, there were no differences in the speed of acceptance, as rhymes, of the two types of nonword. The implications of these results for interactive and autonomous models of spoken word recognition are discussed. It is concluded that the differences in rejection of nonrhyming nonwords are due to the operation of a guessing strategy.

A recurring issue in psycholinguistics is the question of interactivity. There are two good reasons for this. The first is that this topic goes right to the heart of the discipline, with the two sides of the argument making radically different claims about the basic architecture of the language processor. The second is that this is a difficult issue to resolve. In the area of spoken word recognition, both interactive theories, exemplified by the TRACE model (McClelland & Elman, 1986) and the Cohort model (Marslen-Wilson, 1987), and autonomous theories, such as the race model (Cutler, Mehler, Norris, & Segui, 1987) and the fuzzy logical model of speech perception (Masaro & Cohen, 1991), can account for a wide range of data. Cutler et al. (1987) have argued that the power of both classes of model makes it very difficult to find empirical tests to distinguish between them. The present paper is an attempt to provide one such test.

An interactive account of spoken word recognition that has received much recent attention is the TRACE model. McClelland and Elman (1986) present an interactive activation model with three levels of processing: the feature-node level, the phoneme-node level, and the word-node level. Input occurs to the feature nodes, and information

flows bottom-up from this level to the phoneme and word levels. Information also flows top-down, so that later stages of processing influence the operation of earlier stages. The evidence that a particular speech token (feature, phoneme, or word) has been heard is represented by the degree of activation of the node corresponding to that item. There are inhibitory interconnections between nodes within any level of processing, as well as facilitatory connections between nodes at different levels of processing. Word nodes therefore interact with phoneme nodes in a top-down manner: activation of the node for *bat* will boost the activation of the nodes of its constituent phonemes /b/, /æ/, and /t/. All phonetic decisions are based on the output from the phoneme nodes. Thus any lexical effects in phonetic decision tasks must be due to the top-down connections. In phoneme monitoring, for instance, detection of the target phoneme has been found to be faster when the target is in a word than when it is in a nonword (Cutler et al., 1987; Rubin, Turvey, & Van Gelder, 1976). In TRACE, this is because activation of the node for the target-bearing word increases, via top-down facilitation, the activation of the target phoneme node.

I will contrast the TRACE model with an autonomous model of lexical effects on phonetic decisions. The autonomous race model (Cutler et al., 1987; Cutler & Norris, 1979; McQueen, 1991a, 1991b) makes the claim that information flow is strictly bottom-up: there is no top-down interaction between levels of processing. Phonetic decisions can be made via either a prelexical or a lexical route. In the prelexical case, the decision is based on the processing that operates for lexical access. The prelexical procedure races with a lexical procedure, in which phonetic decisions are based on the phonological codes associated with each lexical entry. The race operates on a trial-by-trial basis, so that the route that is faster on any particular trial is responsible for that decision. The lexi-

This research was supported by an MRC Research Studentship, and by Joint Councils Initiative Grant E304/148. It forms part of a doctoral dissertation submitted to the University of Cambridge. Experiments 2 and 3 were reported at the Cambridge meeting of the Experimental Psychology Society in July 1989. I would like to thank my supervisor, Anne Cutler, for her assistance at all stages of this research, and Dennis Norris, for useful discussions and assistance with computer software. I would also like to thank Arthur Samuel and an anonymous reviewer for helpful comments on a previous version of this paper, and the Longman Group U.K. Inc. for access to a machine-readable version of the *Longman Dictionary of Contemporary English*. Correspondence concerning this article should be addressed to J. M. McQueen, MRC Applied Psychology Unit, 15 Chaucer Rd., Cambridge CB2 2EF, England.

cal effect in phoneme monitoring is explained as follows: only responses to targets in words can be made via the lexicon; on a proportion of these trials, the lexical route will win the race, and this will shift the reaction time (RT) distribution of the word responses to a lower mean RT relative to the nonword response distribution, which is due entirely to decisions made via the prelexical route.

Both the TRACE model and the race model can account for many instances of lexical effects on phonetic decision making. They can deal with the variability of lexical effects in phoneme monitoring (see Cutler et al., 1987). Similarly, both can account for lexical involvement in initial-position phonetic categorization (more lexically consistent categorization responses in the ambiguous region of continua such as *dice-tice* and *dype-type*; Connine & Clifton, 1987; Ganong, 1980).

One effect in phonetic categorization, however, is predicted by TRACE, but not by the race model. This is the lexical mediation of compensation for coarticulation, reported by Elman and McClelland (1988). Listeners were more likely to label an ambiguous stop (/ʔ/) between /t/ and /k/ as /k/ in the context "christmas ?apes" (for example), and as /t/ in the context "foolish ?apes." This effect was attributed to a low-level compensation for coarticulation process: the perceptual system compensates for /t/ being more /k/-like (i.e., with a more posterior place of articulation) following an /f/, and for /k/ being more /t/-like (with a more anterior place of articulation) following an /s/. Elman and McClelland demonstrated that this effect also appears when the word-final fricative was ambiguous, midway between /s/ and /f/ in, for example, "christma?" and "fooli?" They argued that the fricative was disambiguated on the basis of lexical knowledge, top-down from the lexicon, and that the "restored" fricative was then able to influence the compensation for coarticulation process. Lexical involvement in a low-level perceptual process, via top-down connections from word to phoneme nodes, is predicted by TRACE. This effect causes problems for the autonomous race model. Note, however, that Norris (in press) has demonstrated this effect in a dynamic connectionist network, where on-line processing is strictly autonomous, as in the race model.

On the other hand, some data from the categorization task cause problems for the interactive account that TRACE gives. McQueen (1991a, 1991b) has shown that for monosyllabic words, TRACE makes RT predictions in word-final categorization that have not been confirmed. In TRACE, top-down facilitation increases over time, as activation builds up on lexical nodes. Thus TRACE predicts that lexical effects in word-final categorization should increase over time, at least over a time period several times the length of a monosyllabic target-bearing word (estimated from McClelland, 1987). In contrast to this prediction, McQueen (1991a, 1991b) found that lexical effects in word-final categorization *decreased* over time. This RT pattern is predicted by the race model. It should

be noted, however, that it remains theoretically possible for interactive theories to explain this effect, under the assumption that activation dies away very rapidly after word offset. But, given its current parameters, this result nevertheless contradicts the TRACE model.

Thus, even within the categorization paradigm, results do not unanimously support either model. This discussion highlights the stalemate that could develop if interactive and autonomous theories were contrasted as general classes. It therefore reinforces the need to focus on the predictions of specific instantiations of the theories for progress to be made. It seems important to continue to pit the predictions of the specific models against each other. This was attempted here in a rhyme-judgment task.

One result that is of particular relevance to the following experiments is the failure to find inhibitory lexical effects in phoneme monitoring for targets in nonwords. Frauenfelder, Segui, and Dijkstra (1990) have shown that there are reliable lexical effects (faster responses to targets in words than in matched nonwords) in monitoring for a phoneme that occurs after the word's uniqueness point (the point at which, as one moves from left to right through the word, there is a unique lexical entry consistent with the input). Thus, responses to /p/ are faster in the Dutch word *olympiade* than in the nonword *arimpiako*. Pitt and Samuel (1991) have shown that similar facilitatory effects may also occur when the target occurs before the uniqueness point. The TRACE explanation for this *facilitatory* effect, that activation of the word node for *olympiade* boosts the activation of the /p/ node, predicts that there will be *inhibitory* lexical effects within nonword decisions. Top-down facilitation from word nodes to constituent phonemes will increase the degree of inhibition, via phoneme-to-phoneme inhibitory connections, on nonconstituent phonemes. Thus detection of a /t/ in the French nonword *vocabulaire* should be inhibited by top-down facilitation of /l/ from the activated *vocabulaire* node. Frauenfelder et al. found no evidence of this lexically mediated inhibition: detection of /t/ in *vocabulaire* was not inhibited, relative to the matched nonword *socabulaire*.

The race model predicts the facilitatory effects in phoneme monitoring for the reason given above: lexical route responses will contribute to word decisions, increasing the mean speed of the response distribution for words relative to that for nonwords. Varying the position of the target phoneme relative to the uniqueness point will alter the likelihood of obtaining a facilitatory lexical effect. The later the target phoneme and the earlier the uniqueness point, the more likely it will be that the lexical route will win the race. But the lexical route can still win the race with a preuniqueness point target, with the resulting lexical advantage. The model thus predicts facilitatory effects for targets both before and after a word's uniqueness point. The race model also claims that all phonetic decisions about nonwords are insulated from lexical effects, because all responses made to nonwords have to be made

via the prelexical route. It thus predicts that there should be no inhibitory effects on nonwords, as Frauenfelder et al. (1990) found.

The TRACE model predicts that lexical involvement should be detectable in phonetic decisions made in response to nonwords because of the assumption of top-down processing. The lexicon influences the prelexical processing of any speech input, irrespective of its lexical status. The race model makes the opposing assumption that the lexicon can only be involved in on-line phonetic decisions to words. This distinction, that TRACE predicts lexical effects on phonetic decisions to nonwords while the race model does not, can be tested in rhyme-decision tasks.

Rhyme monitoring may be a more reliable task than phoneme monitoring for examining lexical involvement in phonetic decision making. A defining feature of lexical influence in phoneme monitoring is variability. Lexical effects come and go, depending on parameters such as task demands and the nature of the stimulus materials (Cutler et al., 1987). For example, these authors found that an RT advantage for words over nonwords is absent when subjects monitor for phonemes in strictly monosyllabic lists. A task in which lexical effects are less dependent on particular features of the experimental situation is likely to be a more effective tool for examining interactive and autonomous accounts of lexical influences. Rhyme monitoring may be such a task simply because it requires decisions to be made about more of a stimulus word than phoneme monitoring. Since subjects can attend to particular locations in stimulus strings in phoneme monitoring (Pitt & Samuel, 1990), they may tend to ignore the lexical status of stimuli in this task. Rhyme decisions involve the vowel and final consonant(s) (the *rime*) of monosyllables. Subjects therefore have to treat stimuli as strings in rhyme monitoring, and they cannot focus their attention on a particular phonemic location. They should thus be more likely to treat stimuli as whole words (where possible) in rhyme monitoring.

EXPERIMENT 1

Are lexical effects more reliable in rhyme monitoring than in phoneme monitoring? Seidenberg and Tanenhaus (1979) have shown that rhyme monitoring is faster when the rhyme target and the preceding cue are orthographically consistent (e.g., *pie-tie*) than when they are dissimilar (e.g., *rye-tie*). Their subjects were presented with a cue word (either visually or auditorily) followed by auditory lists. They were required to press a key when they heard a word that rhymed with the preceding cue. The effect of orthography was replicated by Donnenwerth-Nolan, Tanenhaus, and Seidenberg (1981), again using this rhyme-monitoring task. The presence of the orthographic effect suggests that lexical access (including the accessing of orthographic information) occurs in this rhyme-judgment task. These authors also demonstrated a semantic priming effect: monitoring on targets like *bite* following the cue *kite* was faster when *bite* was preceded

by a semantically related word in the auditory list (e.g., *chew, bite, told*) rather than in a list without a semantic associate (e.g., *vest, bite, told*). This semantic priming indicates that word meanings were available during monitoring, strengthening the claim that the task involves lexical access. The most direct test of lexical involvement in rhyme monitoring, however, would be to compare rhyme decisions to words and nonwords. Experiment 1 was a test of whether or not rhyme decisions are easier for words than for nonwords.

Another way to demonstrate lexical involvement in a phonetic decision task is to show a word-frequency effect (see, e.g., Dupoux & Mehler, 1990). Although there is much debate about the locus of word-frequency effects, particularly in visual word recognition (see, e.g., Monsell, Doyle, & Haggard, 1989), an RT advantage for rhyme decisions to higher frequency words can be explained, at least in part, by assuming that word recognition is easier for words that are encountered frequently than for words that are heard only rarely (see Luce, 1986). Thus, high- and low-frequency words were presented in Experiment 1. Note that the frequency effect is only intended to be an indicator of lexical involvement in the rhyme task. No strong claim is being made that frequency effects in spoken word recognition are due entirely to processes of lexical access. Such effects may of course also be due to both prelexical factors, such as the frequency of occurrence of diphones or triphones (or bigram/trigram frequency), and postlexical processes, such as decision bias. Nonetheless, a frequency effect can at least be taken to be consistent with lexical involvement in rhyme monitoring.

If lexical involvement in rhyme monitoring is reliable, lexical effects should be equivalent in different contexts. In a monitoring task, the proportions of words and nonwords can be varied. The influence of the lexicon can be assumed to be more reliable if it is as large in a comparison of lists containing only words or only nonwords as it is in lists containing a mixture of words and nonwords. Rubin et al. (1976) found that word-nonword differences in a phoneme-monitoring task were larger when mixed lists of words and nonwords were presented than when pure lists of only words or only nonwords were presented. Clearly, lexical effects would be shown to be more robust in rhyme monitoring than in phoneme monitoring if no such difference was found here. Experiment 1 therefore included a comparison of performance on pure lists (only words or only nonwords) and mixed lists.

Method

Materials and Procedure. Rhyme targets were presented in the penultimate position in auditory lists. Each of these lists was preceded by a visual cue word. There were three primary variables, manipulated within subjects: the lexical status of the rhyme targets, the frequency of occurrence of the targets that were words, and the type of list in which the targets appeared. Each subject received a block made up only of lists of words, a block of only nonwords, and two blocks containing lists with a mixture of words and nonwords. The word targets were split into two equal sets of high- and low-frequency words (HF and LF words, respectively), but this

frequency manipulation was not blocked: HF and LF words were randomly presented within the words-only and mixed blocks. All subjects were presented with all of the targets once, and the matched HF and LF word targets were counterbalanced, between subjects, across the words-only and mixed blocks. An additional variable of order of presentation of blocks was manipulated between subjects. Thus one third of the subjects received the words-only block first, one third received the mixed blocks (words and nonwords) first, and the remaining third received the nonwords-only block first.

The critical stimuli consisted of 24 rhyme cues, each of which was yoked to 4 targets. These targets were an HF word, an LF word, and two nonwords (NW). Thus, for example, the probe *dig* was yoked to *big* (HF), *pig* (LF), and *kig* and *chig* (NW). A complete list of these stimuli is given in Appendix A. The sets of HF and LF words were as dissimilar in frequency as possible: the HF words were > 160 counts per million and the LF words were < 25 counts per million in the Francis and Kučera (1982) frequency norms. The cue words had a mid range of frequencies, such that they were lower in frequency than the matched HF target and higher than the LF target.

The lists were presented in blocks of 37. Each block contained 24 lists in which there was a critical rhyme target. These lists varied in length from 4 to 6 items. The remaining 13 lists varied in length from 2 to 6 items. Five of these appeared as the first lists in each block. The remaining 8 were spaced randomly through the block. The last item in every list was always the word *end*. Each critical cue was seen by each subject six times: followed by each of the 4 critical targets, followed once by an additional target (see Appendix A—these targets were required for counterbalancing within the experimental and practice blocks), and followed once by no target at all. No critical cue appeared more than twice in any block. Three practice blocks consisting of 8 lists were also constructed: one with lists of only words, one with mixed words and nonwords, and one with only nonwords. Finally, 442 items (half words, half nonwords) were used as filler items in the lists. The use of filler items in lists of varying length makes target location unpredictable and allowed list composition to be manipulated. All stimuli were monosyllabic.

The stimulus blocks were recorded by the author (a native speaker of British English) in a sound-attenuated booth onto the left channel of a tape with items being spoken at a rate of one per second. There was a pause of 4 sec between each list. In total there were three practice blocks recorded and seven experimental blocks—two words-only blocks, four mixed blocks (thus counterbalancing HF and LF words across blocks), and one nonwords-only block. Timing pulses were added to the right channel of the tape at the onset of each critical target; these were used to start the timer of a microcomputer, which was stopped by the subject pressing a response key. It seemed more appropriate to measure rhyme-monitoring RTs from the vowel onset of the targets, rather than from the onset of the initial phoneme, so the stimuli were digitized, sampling at 10 kHz and using 12-bit A/D conversion, and the waveforms were examined with a speech editor. Measurements were taken between the timing pulse and the vowel onset for each target. These values were subtracted from the relevant RTs for each target for each subject prior to data analysis.

The subjects were tested individually in a quiet room. Auditory materials were presented binaurally, over headphones; visual materials were presented in lowercase on the screen of a microcomputer. The subjects were seated in front of the computer and told that they would see words on the screen and then hear lists of words and nonwords over the headphones. They were asked to listen to the lists and to press a button marked YES if they thought they heard a rhyme of the cue that had just appeared on the screen. They were told that each list ended with the word *end*; they were asked to press a button marked NO if they got to the end of a list without hearing a rhyme. The YES button was held in the preferred hand, the NO button in the other hand. Each subject was given the

practice block appropriate to the intended order of presentation of the experimental blocks (e.g., the mixed practice block before the mixed experimental blocks), one words-only block, two mixed blocks, and the nonwords-only block.

Subjects. The subjects were 24 student volunteers from St. John's College, Cambridge, who were paid for their participation. There were 15 female and 9 male subjects, with no known hearing loss; they were between 18 and 25 years of age.

Results and Discussion

The data that were analyzed were RTs measured from the vowel onset of each target. RTs of less than 200 msec or greater than 1,500 msec were excluded from the analysis. Including errors, the missing data constituted about 6% of the total data set. Analyses of variance (ANOVAs) were carried out on the RT and error data, with subjects and items as random factors. The mean RTs and error rates are shown in Table 1.

The RT analysis showed that word responses were reliably faster than nonword responses, with the 54-msec effect significant by subjects [$F_1(1,18) = 33.27, p < .001$] and by items [$F_2(1,94) = 22.32, p < .001$]. This yielded $\min F'(1,78) = 13.36, p < .01$. There was also a significant effect of frequency, with responses to HF words, on the average, 38 msec faster than responses to LF words [by subjects, $F_1(1,18) = 28.84, p < .001$, and by items, $F_2(1,46) = 6.66, p < .05$, giving $\min F'(1,61) = 5.41, p < .05$]. The effect of list type was not significant in either analysis. There was also a lexical effect in the error data, with more errors made on words than on nonwords. This effect was significant by subjects [$F_1(1,18) = 11.17, p < .005$] and by items [$F_2(1,94) = 6.01, p < .05$], but $\min F'$ was not significant. There were no significant frequency or list-type effects in the error analyses.

In this rhyme-monitoring task, the subjects were reliably faster at detecting rhyming words as opposed to rhyming nonwords. Within the words, there was also a reliable frequency effect: HF words were responded to more rapidly than LF words. These effects were equivalent in pure lists (lists of only words or only nonwords) and mixed lists of words and nonwords. In addition, the subjects made more errors on words than on nonwords. The significant lexical effects found here show that lexical knowledge is used in the rhyme-monitoring task. The significant frequency effect is consistent with lexical involvement, and it serves to corroborate the basic lexical

Table 1
Experiment 1: Mean Reaction Time (in Milliseconds) and Percent Errors for Each Target Type (High-Frequency Word, Low-Frequency Word, or Nonword) in each List Type (Words or Nonwords Only, or Mixed Words and Nonwords)

Target Type	List Type			
	Pure		Mixed	
	RT	% Error	RT	% Error
High-frequency words	520	9	513	6
Low-frequency words	561	6	549	6
Nonwords	588	4	593	4

effect. The failure to find an effect of list type indicates that the lexicon is as involved when the listener is presented with only word stimuli as it is when both words and nonwords are presented. Note that this does not mirror the results from the phoneme-monitoring task, where list structure appears to determine lexical involvement (Cutler et al., 1987; Rubin et al., 1976). This indicates that lexical involvement is a more mandatory feature of rhyme monitoring than of phoneme monitoring, perhaps because the rhyme task involves an analysis of the complete target string. A related argument has been made in the comparison of "standard" and "generalized" phoneme monitoring (Frauenfelder & Segui, 1989).

The TRACE explanation for the lexical and frequency effects found in Experiment 1 is that top-down facilitation increases the level of activation of the appropriate phoneme nodes. Although TRACE does not include a mechanism to account for frequency effects, McClelland and Elman (1986) suggest that a more completely specified model would be one in which higher frequency words have a higher resting level of activation (as in the interactive activation model of visual word recognition; see McClelland & Rumelhart, 1981). With a higher resting level of activation, the facilitatory efficacy of higher frequency words would be greater, and HF words would be responded to more rapidly than LF words, as found here.

The race model also explains the results of Experiment 1 by analogy with the explanation it gives for other lexical effects. Phoneme-monitoring responses can be made on the basis of either a prelexical code or a phonological code associated with a lexical representation. This should also be true for rhyme-monitoring responses. Lexical effects thus reflect decisions being made via the lexical route. Because the two routes race, the proportion of decisions made via the lexicon will be faster, simply because they must have been faster to win the race. Thus, responses to rhyming words will tend to be faster than responses to rhyming nonwords. On this account, the frequency effect is due to the fact that the lexical route is more likely to win the race the higher the frequency of the presented word. A postaccess checking mechanism (see Norris, 1986), necessary for the resolution of lexical ambiguities, may lower the recognition criterion for HF words relative to LF words.

EXPERIMENT 2

The results from Experiment 1 suggest that rhyme tasks may be of use in testing divergent predictions of the TRACE and race models. A variation on the rhyme-monitoring task was therefore used in Experiments 2 and 3 to test the predictions of the models for performance on different types of nonwords. TRACE predicts that lexical involvement should be detectable in rhyme decisions to words and nonwords, since top-down facilitation influences the processing of any input to the phoneme nodes, irrespective of lexical status. Thus, in TRACE, nonwords should be differentially affected by the lexicon, depend-

ing on the similarity of the nonwords to real words. The race model predicts that there should be no lexical effects on rhyme decisions in response to the presentation of nonwords. All perceptual decisions about nonwords must be made via the prelexical route, so these decisions should be insulated from lexical involvement.

Consider stimuli that are phonetic near-neighbors of rhymes. For the rhyme cue *rob*, words such as *shop* and nonwords such as *jop* are phonetic neighbors. They almost rhyme. Stimuli such as these, which will be labeled as *foils*, formed the basis of Experiments 2 and 3. A rhyme-decision task was used, which was a variant on the rhyme-monitoring task. Visual cues were presented, followed by pairs of auditory items. The subjects were required to make a rhyme decision on both members of every pair. This allowed measurement of both "yes" responses to targets and "no" responses to foils.

TRACE predicts that the correct detection of rhyming words should be faster and more accurate than the detection of rhyming nonwords. Such a finding, of course, would be a simple replication of the results of Experiment 1. Top-down facilitation should increase the activation levels of appropriate phoneme nodes. TRACE also predicts that correct rejection of nonrhyming words should be faster and more accurate than rejection of nonrhyming nonwords. So, for example, following the cue *rob*, TRACE predicts that "yes" responses to presentations of the rhyming word target *job* should be faster than those to the presentation of the nonword target *shob*, and that "no" responses to presentations of the nonrhyming word foil *shop* should be faster than those to presentations of the nonword foil *jop*. With the foils and targets interchanged, this pattern should also be found on the same stimuli, following a matched cue, such as *hop* ("yes" responses to *shop* and "no" responses to *job* should be faster than to *jop* and *shob*, respectively).

The race model shares these two predictions: "yes" responses should be faster and more accurate with word targets than with nonword targets, and "no" responses should be faster and more accurate with word foils than with nonword foils. Responses to words can be made via either the prelexical or the lexical route, but nonword stimuli can only be responded to because of the operation of the prelexical route. The lexical route will be successful only for word stimuli, and this will tend to increase the speed of both target ("yes") and foil ("no") word responses.

Now consider two different classes of nonword foils: those that have rhyming lexical neighbors and those that do not. The foil *jop*, following the cue *rob*, has a phonetic neighbor that is a word that rhymes with the cue: *job*. But the foil *vop*, following *rob*, does not have a word neighbor that rhymes; *vob* is also a nonword. Word foils will be designated as W, nonword foils that have a rhyming lexical neighbor, as NW_{+L}, and nonword foils that do not have a rhyming lexical neighbor, as NW_{-L}. Thus, following the cue *rob*, there can be three types of foil: W-*shop*; NW_{+L}-*jop*; and NW_{-L}-*vop*. These labels will

also be used to describe the same items when they follow a matched cue such as *hop*, with which they do rhyme. Here, of course, they are targets.

The TRACE model predicts that top-down facilitation should make it harder to reject NW_{+L} foils (*jop*, following *rob*) than NW_{-L} foils (*vop*). This is because in the former case, the target should at least partially activate a lexical entry (*jop* will activate *job*), and this lexical activation will in turn boost the activation of the appropriate word-final phoneme node (/b/), delaying the detection of the correct phoneme (/p/) because of the phoneme-level inhibitory interconnections. NW_{+L} responses should therefore be slower and/or less accurate than NW_{-L} responses.

The final prediction of TRACE relates to the positive responses to nonword targets. Correct detection of NW_{+L} targets (*jop* following *hop*) should be slower and/or less accurate than correct detection of NW_{-L} targets (*vop*). Again, the NW_{+L} stimulus (*jop*) will partially activate its rhyming lexical neighbor (*job*), and the resulting top-down facilitation to the /b/ phoneme node will inhibit the activation of the /p/ phoneme node.

The TRACE predictions for performance on nonwords are in fact more complex than what has just been described. The recognition of *jop* is not only influenced by the existence of the word *job* in the lexicon. Recognition will depend on the combined influence of *all* the words that are phonetic neighbors of *jop*. In particular, in a task in which all of the items are monosyllabic, rhyme-decision performance will depend on how many words begin with /dʒɒ/. This is the case whether *jop* is a target (after *hop*) or a foil (after *rob*). The combined /dʒɒ/-initial words will all provide top-down facilitation, making it more difficult to recognize *jop*. Thus, TRACE predicts differences between the NW_{+L} items (*jop*) and the NW_{-L} items (*vop*) only to the extent that there are more lexical competitors for the former items than for the latter. This prediction holds for both targets and foils. In keeping with this prediction, the NW_{+L} items had more competitors than did the NW_{-L} items. But it should be noticed that the targets and foils, so defined, differ in one important respect. An NW_{+L} foil includes in its competitor set a word that *rhymes* with the preceding cue; an NW_{-L} foil does not. This distinction does not hold when the same items serve as rhyme targets.

The accounts given above of TRACE and race predictions are based on the way in which the two models process the speech input, and they do not take into consideration other aspects of the rhyme-decision task. Specifically, they do not address the possibility of subjective strategies playing a role in rhyme decision making. In this task situation, where subjects are asked to respond as quickly as possible, it is likely that they might guess that they have heard rhyming targets. After hearing the correct vowel, they might anticipate that the target is a rhyme, before analysis of the final consonant(s) has been completed. The presence of such a guessing strategy can be tested by analyzing the *foil* data. The strategy predicts more false alarms in response to NW_{+L} items (guessing

that the foil *jop* was the target *job* after the cue *rob* without sufficient processing of the final phoneme of *jop*) than to NW_{-L} items (such as *vop*), which do not have rhyming lexical neighbors.

A guessing strategy would also have consequences for RTs. On occasions when the false alarm is not made before the information from the final phoneme becomes available, there will be a decrement in processing speed. Rejection of the incorrect hypothesis *job* will delay the decision that the presented foil *jop* does not rhyme with *rob*. That is, the guessing strategy predicts that correct rejection of NW_{+L} foils will be slower than correct rejection of NW_{-L} foils.

This strategy will therefore produce the RT and error patterns predicted by the TRACE model for performance on *foils*. But it is important to notice that the guessing strategy does *not* predict differences between nonword types in detection of *targets*. Although guessing might serve to increase the difference between word and nonword targets (correctly guessing that *job* has been presented after the cue *rob* should speed W responses relative to NW responses), this strategy does not distinguish between the two classes of nonword target. There is no word rhyme to be guessed given either a NW_{+L} target such as *shob* (after *rob*) or an NW_{-L} target such as *vob*.

The design of and predictions for Experiment 2 can now be summarized. Both models predict differences between words and nonwords. Thus positive and negative rhyme decisions in response to presentations of word and nonword targets and word and nonword foils were compared. The race model predicts no differences due to lexical neighborhood between nonwords. TRACE, however, predicts that nonwords that share onsets with many words will be harder to accept or reject than nonwords that share onsets with few words. Thus, for *both* targets and foils, TRACE predicts a difference between NW_{+L} and NW_{-L} items. The distinction between these nonword items was also motivated by the need to control for guessing strategies in rhyme decision. Listeners may tend to wrongly accept, as rhymes, nonword foils that have rhyming lexical neighbors. Thus, NW_{+L} and NW_{-L} foils differed not only in the number of their lexical competitors, but also according to whether or not this lexical candidate set included a word that actually rhymed with the preceding cue.

In order to provide independent evidence for the operation of a guessing strategy, an additional variable was included in the experiment: the context of each rhyme decision was varied systematically. There were therefore three types of stimuli: targets, foils, and fillers (which, unlike foils, did not share vowels with preceding cue words). All possible pairwise combinations of these three stimulus types were presented, except that there were no foil-foil or filler-filler pairs. Reducing in this way the number of stimulus pairs to which two "no" responses would be expected meant that following any given visual cue, it was more likely that there would be at least one "yes" response. If subjects were to guess, this arrangement of stimuli would be more likely to yield false alarms

than misses. Furthermore, this arrangement would be likely to yield false alarms in response to the second stimulus in each pair when the first stimulus has produced a "no" response. Therefore, a strategy of guessing (based on a response on insufficient analysis of the signal) would predict more false alarms in response to foils after fillers than after targets.

Method

Materials. Rhyme decisions were made in response to presentations of three types of items: targets, where the item presented rhymed with the preceding cue word; foils, where the item presented did not rhyme with the cue, but was a near neighbor of the cue, sharing the same vowel but differing in the nature of the final consonant; and fillers, where the item presented did not rhyme with the cue and was not a near neighbor of the cue, differing in both vowel and final consonant. The cues, targets, and foils were matched so that the stimuli served as both targets and foils according to the cue presented. All the stimuli were monosyllabic. The targets and foils are shown in Appendix B.

The stimuli were constructed from a base list of 18 matched pairs of cue words. For each cue, three targets were constructed, one word and two nonwords. The three targets for one member of a given cue pair also served as foils for the other cue. The three target/foil types, W, NW_{+L}, and NW_{-L}, have been described above. The number of monosyllabic words beginning with the same consonant and vowel was counted for each item, using a machine-readable version of the *Longman Dictionary of Contemporary English*. The mean number of lexical competitors in each condition were: W, 7.9; NW_{+L}, 7.7; and NW_{-L}, 5.3. Note that the items that served as target words following a given cue have the same competitor environment as those that served as NW_{+L} foils, and vice versa following the matched cue. Pairwise comparisons were carried out on the number of competitors. There was a significant difference between the W and NW_{-L} items [$t(17) = 1.78, p < .05$, one-tailed], as well as a significant difference between the NW_{+L} and the NW_{-L} items [$t(17) = 2.11, p < .05$, one-tailed]. There was no significant difference between the NW_{+L} and the W items. There were therefore reliably more lexical competitors in the W and NW_{+L} conditions than in the NW_{-L} condition. In addition, when the NW_{+L} items served as foils, one of the lexical competitors was a word that rhymed with the preceding cue (e.g., *job* given *jop* after the cue *rob*). However, there was no rhyming word given a NW_{-L} foil (e.g., *vob* given *vop* after the cue *rob*).

The experiment was split into two blocks so that all subjects received each of the 36 cue words twice, once in each block. The subjects were split into three groups, each receiving 72 of the total 108 targets and 36 of the total 108 foils. These were balanced across the cue items so that each subject received 24 of each of the three target types and 12 each of the W, NW_{+L}, and NW_{-L} foils. Each subject also received 36 filler items, so that error free performance would result in an equal number of "yes" and "no" responses. The total number of words and nonwords were balanced in each set. Each cue was followed by two decision words, a response being required to both. This manipulation allowed the context of a given response to be systematically varied. Targets could appear in the context of another target, a foil, or a filler, and foils in the context of a target or a filler.

The cues were presented visually, in lowercase on a micro-computer screen, and the auditory items were presented binaurally over headphones. The stimuli were recorded by the author in a sound-attenuated booth, onto the left channel of three tapes, one for each group of subjects. Timing pulses (inaudible to the subject) were placed on the right channel of the tape, aligned approximately with stimulus onset. Prior to analysis, each stimulus item was digitized at a sampling rate of 20 kHz, using 12-bit A/D con-

version, and the time between the timing pulse and the onset of periodic energy corresponding to vowel onset was measured on each waveform. RT could thus be measured from vowel onset.

Procedure. The subjects were seated in a small, quiet room and instructed that they were taking part in a rhyme-detection task. They were asked to respond as quickly and as accurately as possible to every word they heard, deciding whether it rhymed with a preceding visual cue presented on a computer screen. It was made clear that each visual cue would be followed by two decision items and that subjects were expected to respond to the first member of each pair before hearing the second.

In all cases, the visual cue was presented for 1.5 sec. After a 1.5-sec silence, the first auditory item was presented. The onset of the second auditory item fell 2 sec after the onset of the first. There were a further 2.5 sec after this second onset prior to the next visual cue. All subjects received a short practice block, consisting of 8 filler cues and 16 auditory decision items. Then the two experimental blocks were run. Each block consisted of 40 visual cues and 80 rhyme decision responses. In each of the blocks, there were first four warm-up trials prior to the experimental trials. Thus, overall, there were 144 responses per subject. The microcomputer controlled the timing of the stimulus presentations and stored the subjects' RTs. The subjects' responses were made on two buttons labeled YES and NO, the preferred hand being used to make YES responses.

Subjects. Thirty members of the Medical Research Council Applied Psychology Unit subject panel were paid to take part in the experiment. There were 18 females and 12 males, with an age range of 20 to 47 years. No subject reported any hearing loss.

Results and Discussion

Throughout this section, *rob* will be used as an example cue word, followed by the targets *job*, *shob*, and *vob*, and the foils *jop*, *shop*, and *vop*. First, the target responses were analyzed. The mean RTs and error rates are shown in Table 2. In the RTs there were significant differences among the three conditions, as measured with two ANOVAs, once by subjects [$F_1(2,54) = 39.6, p < .001$], and once by items [$F_2(2,51) = 7.1, p < .001$]. This yielded a significant min $F'(2,69) = 6.04, p < .01$. Planned comparisons indicated that this effect was due entirely to differences between words and nonwords. The mean difference of 72 msec between words (W - *job*) and nonwords with more lexical neighbors (NW_{+L} - *shob*) was significant [by subjects, $t(29) =$

Table 2
Experiments 2 and 3: Mean Reaction Time (in Milliseconds) and Percent Errors for Detection of Rhyme Targets of Three Types, W, NW_{+L}, and NW_{-L}, and for Rejection of Nonrhyming Foils of the same Three Types

Response Type	Stimulus Type					
	W		NW _{+L}		NW _{-L}	
	RT	% Error	RT	% Error	RT	% Error
Experiment 2						
Detection*	733	8	805	12	807	9
Rejection†	810	3	855	8	836	6
Experiment 3						
Detection*	563	5	616	6	610	7
Rejection†	736	16	765	24	748	20

Note—Percent errors are misses for targets and false alarms for foils.
*Targets. †Foils.

-8.08, $p < .001$, and by items, $t(71) = -4.61, p < .001$]. The mean difference of 74 msec between words (*W-job*) and nonwords with fewer lexical neighbors (*NW_{-L}-vob*) was also significant [by subjects, $t(29) = -6.94, p < .001$, and by items, $t(71) = -4.70, p < .001$]. The difference (2 msec, on the average) between the two nonword types was not significant by subjects or by items. Lexical effects appear to be present in the positive rhyme responses to words, but not in the positive rhyme responses to nonwords. An analysis of the errors on targets (misses) gave a significant effect among the three conditions only by subjects [$F_1(2,54) = 4.6, p < .05$].

Table 2 also gives the mean RTs of correct rejections of foils. In an ANOVA by subjects, there was an overall main effect of lexicality [$F_1(2,54) = 12.12, p < .001$], but this was not reliable across items [$F_2(2,105) = 2.16, n.s.$]; min F' was not significant. A planned test based on the subjects analysis showed that this main effect was principally due to two effects: the mean difference of 45 msec between *W* and *NW_{+L}* stimuli [*shop* and *jop*: $t(29) = -4.45, p < .001$], and the average 26-msec difference between *W* and *NW_{-L}* stimuli [*shop* and *vop*: $t(29) = -2.55, p < .05$]. The average 19-msec difference between *NW_{+L}* and *NW_{-L}* stimuli was marginally significant [*jop* and *vop*: $t(29) = 1.85, .05 < p < .1$]. The false alarms in response to foils are also shown in Table 2. These data gave no significant differences on Wilcoxon matched pairs tests, over subjects, for all three comparisons of foil types.

The presence of the guessing strategy was tested in the final analysis, which focused on false alarms. The effects of order of presentation of auditory stimuli and of the context of a foil decision are shown in Table 3. In an ANOVA on the means across subjects, there was a main effect of foil context [targets vs. fillers, $F(1,27) = 4.4, p < .05$], but no main effect of order (foil first vs. foil second). There was, however, a highly significant interaction of these two factors [$F(1,27) = 17.8, p < .001$]. The only differences that were significant in t tests across subjects were those between false alarms in responses to foils following fillers (Ff), and each of the other three conditions: foils before targets (fT) and Ff [$t(29) = -2.6, p < .05$]; foils after targets (Tf) and Ff [$t(29) = -4.3, p < .001$];

Table 4
Experiments 2 and 3: Overall Mean Reaction Times (in Milliseconds) and Percent Errors for Correct Detection of Target Stimuli and Correct Rejection of Foil and Filler Stimuli

Experiment	Targets		Foils		Fillers	
	RT	% Error	RT	% Error	RT	% Error
2	782	10	834	6	692	1
3	596	6	750	20	597	2

and foils before fillers (fF) and Ff [$t(29) = -3.5, p < .005$]. These results therefore suggest that there were more false alarms with foils immediately following a filler—that is, guessing was partially determined by the contingencies that made it more likely that there would be at least one rhyme target in every pair of auditory stimuli than that there would be no target.

It is worth noting that the foil stimuli were reliably more difficult to process than the filler stimuli, as can be seen in Table 4. There were more errors on the foils than on the fillers. In an ANOVA to compare the mean RTs of correct rejections of foils and fillers, filler responses were significantly faster [$F(1,27) = 298.3, p < .001$].

These results suggest that although responses to words (both targets and foils) are facilitated in relation to responses to nonwords, as is predicted by both models, there is no consistent evidence that performance on *NW_{+L}* stimuli is impaired relative to performance on *NW_{-L}* stimuli. There is some evidence that *NW_{+L}* foils are rejected more slowly than *NW_{-L}* foils. There is also evidence that subjects are employing a guessing strategy. After a “no” response, listeners were more likely to false alarm on a foil.

Consider, however, the small number of false alarms in Experiment 2. Given that there were only 60 data points in total, it is difficult to interpret these results. The failure to find significant differences between *NW_{+L}* and *NW_{-L}* false alarms may be a consequence of the small data set. Furthermore, the differences found in the order and context analysis could be artifactual. It was therefore important to attempt to replicate Experiment 2.

EXPERIMENT 3

In this experiment, an attempt was made to increase the proportion of errors on the foil stimuli. The subjects were encouraged to respond as quickly as possible. It was stressed in the instructions that the speed of responding was of primary importance. In addition, the subjects were given feedback after every block of 10 trials which told them how much faster or slower they had been on that block. It was hoped that increasing time pressure would increase the false alarm rate. The aim of the experiment was to replicate the findings of Experiment 2 by using a procedure that differed only in the inclusion of the time-pressure manipulation.

Method

This experiment was identical to Experiment 2 in almost all respects. It differed in that the visual cue words were presented with a Tektronix oscilloscope. Stimulus presentation and response timing were controlled by a PDP 11-23 minicomputer. These minor

Table 3
Experiments 2 and 3: Percent of False Alarms (FA) to Foils, According to the Context of the Decision

Foil Context	Foil Position			
	1st		2nd	
	% FA	Context	% FA	Context
Experiment 2				
Target (T)	5.9	fT	3.0	Tf
Filler (F)	3.3	fF	10.0	Ff
Experiment 3				
Target (T)	18.5	fT	15.7	Tf
Filler (F)	15.3	fF	31.0	Ff

Note—Foils (f) appeared before (fT) or after (Tf) target stimuli, or before (fF) or after (Ff) filler stimuli.

changes were made to facilitate the important change: subjects were presented with feedback messages on the screen of the scope after every 10 trials. Each message was of the form: "X% faster/slower," where X was the change in mean RT for that block of 10 trials relative to the preceding block. The subjects were told that they should respond as fast as they could throughout the experiment, and that they should attempt to increase their speed as they went along. That is, they were asked to try to make the message be "X% faster," with X as large as possible. As in Experiment 2, there were three groups of subjects, who each received one third of the critical materials.

Twenty-four members of the Medical Research Council Applied Psychology Unit subject panel were paid to take part. There were 11 female and 13 male subjects, with no known hearing loss; they were between 17 and 40 years of age.

Results and Discussion

The first aspect of the data to note is that the speed instructions and feedback messages had the desired effect, as can be seen in Table 4. The mean RTs were faster, and there was an increase in the number of errors in response to foils, with little change in errors on fillers and a decrease in errors on targets. In an ANOVA of correct rejection of foils and fillers, filler responses were significantly faster than foil responses [$F(1,21) = 222.1, p < .001$]. The high error rate on foils, compared with more or less equivalent error rates in Experiments 2 and 3 on targets and fillers, suggests that the speed instructions selectively impaired performance on the foil stimuli.

The mean RTs for correct responses to targets are shown in Table 2. Two ANOVAs were performed, with subjects and items as random factors. Significant differences were found among the three conditions [by subjects, $F_1(2,42) = 27.1, p < .001$; and by items, $F_2(2,51) = 3.31, p < .05$], but this yielded a nonsignificant min $F'(2,63) = 2.95, p > .05$. Planned comparisons showed that this effect was due entirely to the differences between word (W) stimuli and nonword (NW_{+L} and NW_{-L}) stimuli: W and NW_{+L} (*job* and *shob*) [by subjects, $t(23) = -7.39, p < .001$; and by items, $t(71) = -3.98, p < .001$]; W and NW_{-L} (*job* and *vob*) [by subjects, $t(23) = -5.32, p < .001$; and by items, $t(71) = -3.15, p < .005$]. There were no significant differences between the nonword types in either analysis. Replicating Experiment 2, there seems to be no evidence here that NW_{+L} performance is impaired in relation to NW_{-L} performance. It is also worth noting that collapsing across the nonword types [by subjects, $F_1(1,21) = 42.5, p < .001$; by items, $F_2(1,34) = 5.6, p < .05$], did yield a significant min $F'(1,42) = 4.9, p < .05$.

An analysis was also performed with the data from Experiments 2 and 3 combined. In an ANOVA, there was a highly significant lexical effect [$F_1(2,96) = 62.04, p < .001$; $F_2(2,51) = 5.85, p < .01$; min $F'(2,61) = 5.35, p < .05$]. Planned comparisons showed that this effect was due to the average 64-msec difference between W and NW_{+L} responses [$t(53) = -10.62, p < .001$, by subjects; $t(71) = -4.80, p < .001$, by items] and the average 63-msec difference between W and NW_{-L} responses [$t(53) = -8.52, p < .001$, by subjects; $t(71) = -4.44, p < .001$, by items]. The mean difference of

1 msec between nonword types was not significant in either analysis.

Table 2 also shows the mean RTs for correct rejection of the three types of foil. Two ANOVAs indicated that there was a marginally significant effect across the three stimulus conditions [by subjects, $F_1(2,42) = 3.5, p < .05$]; but the effect was not reliable across items, and min F' was not significant. Planned comparisons within the subjects analysis showed that this marginal effect was due to faster responses to word stimuli, the difference between W and NW_{+L} items (*shop* and *jop*) being the only significant one [$t(23) = -2.08, p < .05$]. This analysis replicates the results in Experiment 2, in that there were no reliable differences between NW_{+L} and NW_{-L} (*jop* and *vop*) responses. But there was again a tendency for NW_{+L} responses to be slower than NW_{-L} responses (17 msec, on the average).

This analysis was repeated, with the two experiments combined; in an ANOVA, the lexical effect was significant by subjects [$F_1(2,96) = 13.58, p < .001$], but not by items, and min F' was not significant. Planned comparisons based on the subjects analysis showed that all three comparisons were significant: the average 38-msec difference between W and NW_{+L} responses [$t(53) = -4.55, p < .001$]; the average 20-msec difference between W and NW_{-L} responses [$t(53) = -2.60, p < .05$]; and the average 18-msec difference between NW_{+L} and NW_{-L} responses [$t(53) = 2.19, p < .05$]. The tendency to take longer to reject nonwords with more phonetically close lexical neighbors than to reject nonwords with fewer such neighbors was therefore reliable over 54 subjects.

The miss and false alarm data for Experiment 3 are also shown in Table 2. There were no significant effects in the misses. For the false alarms, although there were no reliable differences on Wilcoxon matched pairs tests, over subjects, for two comparisons of foil types, the comparison of W and NW_{+L} responses was significant [$T(23) = 65, p < .05$]. The false alarm data were also collapsed over both experiments. One-tailed Wilcoxon matched pairs tests, over subjects, showed that the difference between W and NW_{+L} foils was reliable [$T(37) = 173, p < .01$]. The difference between W foils and NW_{-L} foils was also significant [$T(31) = 161, p < .05$], as was the difference between the two nonword types [$T(34) = 196, p < .05$]. This last result indicates that the increased number of false alarms in response to NW_{+L} foils as opposed to NW_{-L} foils is reliable.

The data for the order and context effect analysis are given in Table 4. A similar pattern to that in Experiment 2 was found in an ANOVA (on means across subjects). There was a main effect of order (foil first vs. second) [$F(1,21) = 5.0, p < .05$], but the main effect of context (target vs. filler) was not significant. There was, however, a large interaction of these factors [$F(1,21) = 22.9, p < .001$]. The reliable differences were as follows: between foils after fillers (Ff) and foils before targets (fT), $t(23) = -3.1, p < .01$; between Ff and foils after targets (Tf), $t(23) = -3.5, p < .005$; and between Ff and foils before fillers (fF), $t(23) = -5.7, p < .001$.

The findings of Experiment 3 replicate the main findings of Experiment 2, thus allowing stronger claims to be made from the false alarm data. Overall, there appears to be a reliable lexical effect in these rhyme-decision tasks, equivalent to that found in the monitoring task reported in Experiment 1. Rhyme decisions (both positive and negative) are faster in response to presentations of word stimuli as opposed to nonword stimuli. However, NW_{+L} targets (those that have more lexical neighbors) are not slower to process than NW_{-L} targets (with fewer lexical neighbors). This pattern was found in both Experiment 2 and Experiment 3.

The picture is different when we turn to the foil results. Analysis of the combined data revealed a statistically robust increase in the number of false alarms to NW_{+L} foils (*jop*) relative to NW_{-L} foils (*vop*). The speed of rejection of foils showed a similar pattern. The subjects were slower in rejecting NW_{+L} foils (*jop*) than they were in rejecting NW_{-L} foils (*vop*) in both experiments, yielding a marginally significant effect in Experiment 2. This effect was reliable in the combined analysis. The subjects found it more difficult to reject nonwords, as not rhyming, when those nonwords had more lexical neighbors rather than fewer lexical neighbors. But note that in the foil analysis, the NW_{+L} and NW_{-L} items differed in a second way; in the former case, there existed a word that did rhyme with the preceding cue.

Is the difference between NW_{+L} and NW_{-L} foils due to the existence of a particular lexical entry (a rhyming word), or due to the number of lexical competitors? Correlational analyses were performed to compare the number of monosyllabic lexical competitors (taken from the *Longman Dictionary of Contemporary English*) and the mean RT (collapsed across both experiments) for each foil in each condition. There were no reliable effects. Spearman's r was small but positive for word foils [$r(35) = 0.19, p > .1$], and small and negative for nonword foils [$NW_{+L}, r(35) = -0.12, p > .1$; $NW_{-L}, r(35) = -0.02, p > .1$]. Similar analyses were performed using the mean RTs for the same items when they served as targets. It was again found that competitor environment failed to predict rhyme-decision performance. All r s were nonsignificant and negative. These analyses, coupled with the asymmetry in the NW_{+L}/NW_{-L} effect (absent in target performance, present in foil performance), indicate that the foil effect may be due to the specific availability of a rhyming word, rather than the total number of lexical competitors. The foil results therefore suggest, in keeping with the results from the contextual analysis, that the subjects tended to guess. In particular, it appears that they guessed that they had heard a rhyming word when such a word was available—that is, when they were presented with NW_{+L} foils.

GENERAL DISCUSSION

Rhyme decisions appear to be influenced by the lexical status of the target string. All three experiments show

a basic lexical effect: rhyme detection is faster with words than with nonwords. Furthermore, Experiment 1 demonstrated that responses to HF word rhymes are faster than those to LF words. This result is consistent with the claim that the lexicon is involved on line in making rhyme decisions in response to presentations of words.

It can also be concluded that lexical involvement is more robust in rhyme judgment than in phoneme monitoring. Unlike the results of the latter task (see Cutler et al., 1987), lexical effects were found in lists of only monosyllabic items in all three rhyme experiments reported here. The size of the lexical effect did not appear to depend on the relative proportions of words and nonwords presented, contrary to the list-type effect reported for phoneme monitoring by Rubin et al. (1976). Finally, as Experiments 2 and 3 showed, nonrhyming words were rejected more quickly than nonrhyming nonwords, and there were fewer errors on nonrhyming words than on nonrhyming nonwords. This indicates that the lexical effect is reliable for both positive and negative rhyme decisions.

Reliable lexical involvement in rhyme decisions suggests that the task may be an appropriate tool for examining predictions of interactive and autonomous models of speech perception. These predictions have to be considered in the light of an additional feature of rhyme-judgment performance, namely, that subjects have a tendency to guess. In Experiments 2 and 3, the contextual analyses suggested that subjects were using a guessing strategy. Especially in Experiment 3, under time pressure, the subjects showed a tendency toward false alarms on foils, guessing that a rhyme target had been presented. This tendency was most marked for NW_{+L} foils—that is, foils that have lexical neighbors that rhyme with the visual cue. The subjects guessed that *jop* was *job*, a word that rhymed with the preceding cue *rob*. In addition to increasing the false alarm rate, this strategy acted to slow down correct rejection of foils (again, particularly NW_{+L} foils), since on the occasions where a false alarm did not occur, rejection of the incorrect hypothesis (*job*) interfered with the decision that *jop* did not rhyme with *rob*.

What, then, can be concluded about interactive and autonomous models on the basis of these results? The TRACE and race models differ only in the predictions that they make for nonword performance. The results partially support the TRACE model: although there were no differences between the two types of nonword target, there were reliable effects in the nonword foils. NW_{+L} foils were rejected more slowly, and with a higher false alarm rate, than were the NW_{-L} foils. The TRACE account of this effect is that top-down facilitation of phoneme nodes from words sharing the initial consonant and vowel of the foils, and that this inhibitory effect is larger than it is with NW_{-L} foils, which have fewer lexical competitors.

But this explanation is unsatisfactory, for two reasons. First, there is the asymmetry between the target and foil results. If inhibition resulting from top-down facilitation from a set of activated word nodes is responsible for the

differences between NW_{+L} and NW_{-L} foils, there should be equivalent differences between NW_{+L} and NW_{-L} targets. The interactive explanation for the nonword difference does not distinguish between positive and negative rhyme decisions. Second, the correlational analyses of RT and number of competitors indicated that lexical competitor set size did not predict rhyme-decision performance, for either positive (target) or negative (foil) decisions.

It could be argued that the overall differences in set size between NW_{+L} and NW_{-L} items were small, and that TRACE would therefore predict the resulting competition effects to be very small. Although the lack of any correlation with competitor set size goes against this claim, this null effect should be treated with caution. A stronger test of competitor environment effects, perhaps taking word frequency and acoustic-phonetic similarity into account, could produce reliable results in support of the TRACE model. But possible reasons for the null effect in the targets do not explain the presence of an effect in the foils. The most problematic aspect of the data for the TRACE model is therefore the asymmetry between the target and foil results.

These data, however, could be accounted for by a modified version of the TRACE model. This explanation is based on the assumption that the small differences in competitor environment used in this experiment were indeed too small for top-down lexical competition effects to be measurable. The differences in the foil responses would be due to an unusually high degree of top-down facilitation from the rhyming word given an NW_{+L} foil (activation of the /b/-node by *job* interfering with the detection of /p/ in *jop*). Seeing a rhyme cue could prime (boost the activation of) words that rhymed with that cue, producing this large degree of top-down facilitation. This process would produce the desired pattern of performance for both foils and targets, since for targets, prior activation of rhyming words would serve to increase the word-nonword difference, not to produce a NW_{+L}/NW_{-L} difference. This is simply the instantiation of the guessing strategy in TRACE. In order to capture performance accurately, the model needs to include a strategic component, a mechanism whereby the process of spoken word recognition is modified by the demands of the task environment (see Cutler et al., 1987, for a similar argument based on phoneme-monitoring results).

The race model predicts that word responses will be faster than nonword responses, as found. It predicts no differences between nonwords, but the failure to find differences between nonword targets results in the assertion of the null hypothesis, so these data cannot be taken as strong support for the model. The differences between nonword foils challenge the model, since it predicts that nonword decisions should be insulated from lexical involvement. However, the difference between NW_{+L} and NW_{-L} foils was predicted by the guessing strategy, and the contextual analysis provides some independent support for this strategy. Since the data are thus consistent with the joint predictions of the race model and the guessing strategy, they can also be accounted for by the model

if it too is modified to include a strategic component. The guessing strategy can be incorporated in the race model as a process that lowers the recognition criterion for words that rhyme with the preceding cue. Lexical route outputs would become available given NW_{+L} foils, increasing the likelihood of false alarms, and interfering with correct rejections.

The present study therefore indicates that an account of guessing behavior needs to be incorporated into both interactive and autonomous models of spoken word recognition. The suggestion is that both models need to account for the way in which the human listener strategically responds to the demands of a laboratory task. It should be noted, however, that Cutler et al. (1987) have argued that such adjustments can be made more parsimoniously to an autonomous than to an interactive model.

In conclusion, the three experiments reported here show that the lexicon is reliably involved in rhyme judgments about words. Rhyme decisions are faster in response to presentations of words as opposed to nonwords. This involvement appears to be more reliable than that found in other tasks, such as phoneme monitoring and phonetic categorization. Experiments 2 and 3 show that lexical knowledge does not influence positive rhyme decisions about nonwords. However, they also show that lexical knowledge is involved in the rejection of nonrhyming nonwords. A nonrhyming nonword with a phonetic neighbor that is a word that *does* rhyme is at a disadvantage. Listeners appear to guess that they have heard this rhyming word. The present experiments demonstrate the need to include subject strategies in models of spoken word recognition, be they interactive or autonomous. Without the inclusion of strategic mechanisms, no model will accurately capture the nature of spoken word recognition.

REFERENCES

- CONNINE, C. M., & CLIFTON, C. (1987). Interactive use of lexical information in speech perception. *Journal of Experimental Psychology: Human Perception & Performance*, *13*, 291-299.
- CUTLER, A., MEHLER, J., NORRIS, D., & SEGUI, J. (1987). Phoneme identification and the lexicon. *Cognitive Psychology*, *19*, 141-177.
- CUTLER, A., & NORRIS, D. (1979). Monitoring sentence comprehension. In W. E. Cooper & E. C. T. Walker (Eds.), *Sentence processing: Psycholinguistic studies presented to Merrill Garrett* (pp. 113-134). Hillsdale, NJ: Erlbaum.
- DONNENWERTH-NOLAN, S., TANENHAUS, M. K., & SEIDENBERG, M. S. (1981). Multiple code activation in word recognition: Evidence from rhyme monitoring. *Journal of Experimental Psychology: Human Learning & Memory*, *7*, 170-180.
- DUPOUX, E., & MEHLER, J. (1990). Monitoring the lexicon with normal and compressed speech: Frequency effects and the prelexical code. *Journal of Memory & Language*, *29*, 316-335.
- ELMAN, J. L., & MCCLELLAND, J. L. (1988). Cognitive penetration of the mechanisms of perception: Compensation for coarticulation of lexically restored phonemes. *Journal of Memory & Language*, *27*, 143-165.
- FRANCIS, W. N., & KUČERA, H. (1982). *Frequency analysis of English usage: Lexicon and grammar*. Boston, MA: Houghton Mifflin.
- FRAUENFELDER, U. H., & SEGUI, J. (1989). Phoneme monitoring and lexical processing: Evidence for associative context effects. *Memory & Cognition*, *17*, 134-140.
- FRAUENFELDER, U. H., SEGUI, J., & DIJKSTRA, T. (1990). Lexical effects in phonemic processing: Facilitatory or inhibitory? *Journal of*

- Experimental Psychology: Human Perception & Performance*, **16**, 77-91.
- GANONG, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception & Performance*, **6**, 110-125.
- LUCE, P. A. (1986). Neighborhoods of words in the mental lexicon. In *Research on speech perception* (Tech. Rep. No. 6). Bloomington, IN: Indiana University, Speech Research Laboratory, Department of Psychology.
- MARSLÉN-WILSON, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition*, **25**, 71-102.
- MASSARO, D. W., & COHEN, M. M. (1991). Integration versus interactive activation: The joint influence of stimulus and context in perception. *Cognitive Psychology*, **23**, 558-614.
- MCCLELLAND, J. L. (1987). The case for interactionism in language processing. In M. Coltheart (Ed.), *Attention and performance XII: The psychology of reading* (pp. 3-36). Hillsdale, NJ: Erlbaum.
- MCCLELLAND, J. L., & ELMAN, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, **18**, 1-86.
- MCCLELLAND, J. L., & RUMELHART, D. E. (1981). An interactive activation model of context effects in letter perception: Pt. 1. An account of basic findings. *Psychological Review*, **88**, 375-407.
- MCQUEEN, J. M. (1991a). The influence of the lexicon on phonetic categorization: Stimulus quality in word-final ambiguity. *Journal of Experimental Psychology: Human Perception & Performance*, **17**, 433-443.
- MCQUEEN, J. M. (1991b). *Phonetic decisions and their relationship to the lexicon*. Unpublished doctoral dissertation, University of Cambridge.
- 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.
- NORRIS, D. G. (1986). Word recognition: Context effects without priming. *Cognition*, **22**, 93-136.
- NORRIS, D. G. (in press). Bottom-up connectionist models of interaction. In R. Shillcock & G. T. M. Altmann (Eds.), *Cognitive models of speech processing: Sperlonga II*. Cambridge, MA: MIT Press.
- PITT, M. A., & SAMUEL, A. G. (1990). Attentional allocation during speech perception: How fine is the focus? *Journal of Memory & Language*, **29**, 611-632.
- PITT, M. A., & SAMUEL, A. G. (1991, November). *Is auditory word recognition serial or interactive?* Paper presented at the meeting of the Psychonomic Society, San Francisco.
- RUBIN, P., TURVEY, M. T., & VAN GELDER, P. (1976). Initial phonemes are detected faster in spoken words than in spoken nonwords. *Perception & Psychophysics*, **19**, 394-398.
- SEIDENBERG, M. S., & TANENHAUS, M. K. (1979). Orthographic effects on rhyming. *Journal of Experimental Psychology: Human Learning & Memory*, **5**, 546-554.

APPENDIX A
Matched Target Stimuli, Experiment 1

Cue	Critical Targets				Additional Target(s)
	HF Word	LF Word	Nonwords		
1. dig	big	pig	kig	chig	lig/fig
2. pan	man	ban	gan	lan	zan/van
3. dime	time	lime	nime	bime	pime
4. mill	will	rill	vill	thill	pill
5. wife	life	rife	dife	bife	strife
6. sing	thing	ling	shing	ging	ring
7. sand	hand	gland	gand	dand	land
8. sink	think	chink	shink	bink	nink
9. bell	tell	dell	pell	kell	sell
10. fame	same	shame	thame	pame	wame/game
11. song	long	gong	vong	mong	nong
12. pack	back	tack	dack	fack	sack
13. deep	keep	peep	teep	geep	veep/weep
14. hide	side	bide	kide	mide	nide
15. lean	mean	wean	nean	fean	thean
16. like	bike	pike	gike	kike	hike
17. yield	field	shield	thield	tield	lield
18. preach	reach	peach	weach	deach	steach
19. bit	sit	chit	thit	plit	vit
20. page	stage	sage	lage	mage	fage
21. file	mile	rile	jile	thile	shile
22. pub	club	blub	lub	fub	rub
23. vote	note	dote	sote	pote	brote
24. rope	hope	lope	bope	gope	dope

Note—Each cue was yoked to a high-frequency (HF) word, a low-frequency (LF) word, two nonwords, and an extra filler.

APPENDIX B
Experiments 2 and 3: Experimental Materials

Cues	Targets	Foils
	W; NW _{+L} ; NW _{-L}	NW _{+L} ; W; NW _{-L}
1. nib/nip	fib; shib; thib	fip; ship; thip
2. curb/slurp	herb; cherb; lerb	herp; chirp; lerp
3. rob/hop	job; shob; vob	jop; shop; vop
4. pig/chick	jig; thig; shig	jick; thick; shick
5. peg/deck	leg; neg; theg	leck; neck; theck
6. log/clock	fog; sog; thog	fock; sock; thock
7. curd/curt	third; dird; chird	thirt; dirt; chirt
8. bud/rut	thud; shud; lud	thut; shut; lut
9. bard/cart	lard; dard; thard	lart; dart; thart
10. lathe/rage	bathe; pathe; mathe	bage; page; mage
11. chide/brine	bide; pide; jide	bine; pine; jine
12. lone/coach	bone; pone; thone	boach; poach; thoach
13. lease/kneel	geese; keese; theese	geel; keel; theel
14. mirth/purse	girth; kirth; sirth	gurse; curse; surse
15. mess/hen	guess; kess; shess	gen; ken; shen
16. mice/pipe	dice; tice; chice	dipe; type; chipe
17. mash/fang	dash; tash; thash	dang; tang; thang
18. bish/biff	dish; tish; sish	diff; tiff; siff

Note—Targets and foils served respectively as targets and foils following the first cue in each pair, and respectively as foils and targets following the second cue.

(Manuscript received November 9, 1991;
revision accepted for publication August 4, 1992.)