

Episodic effects on the visual comparison of letters in words

HARVEY H. C. MARMUREK
University of Guelph, Guelph, Ontario, Canada

In two experiments, subjects compared the identity of either the first letters of displays or all the letters in displays. Words yielded faster decisions than did pseudowords (pronounceable items derived by changing a vowel in a word) only when entire displays were compared. In a training phase run before the comparison task, subjects had written down either the first letters or entire displays. Relative to items that had not appeared in the training phase, prior experience at writing entire displays facilitated *same* decisions about entire displays and interfered with *different* decisions about entire displays. Prior experiences did not moderate the processing of first-letter identities. It was concluded that episodic effects arise from the interaction of higher order information underlying lexical encoding. Prelexical processing of letters is not sensitive to episodic effects.

The issue addressed in the present experiments concerns the relationship between word units and letter units in visual word perception. Several different views on that relationship have been proposed. James (1892/1961) interpreted the finding of Cattell (a word could be reported at a shorter exposure duration than could letters not forming a word) as evidence that "a word is a conceptual system in which the letters do not enter consciousness separately, as they do when apprehended alone" (p. 86). That interpretation was echoed by Huey (1908), who stated that "there is a hierarchy of recognition habits, the exercise of the higher drafting away the consciousness that would otherwise serve for completing the recognition of the particular letters" (p. 112).

For both James and Huey, the fate of letters is governed by their context such that if letters form a word, then the word's identity, rather than the letters' identities, gains priority to consciousness. Recent information-processing approaches to word perception have explored this idea more analytically. For example, Healy and her colleagues (Healy, Oliver, & McNamara, 1987) have proposed a unitization model of word perception. According to that model, the processing of letter units and the processing of units larger than the letter (e.g., words) are activated in parallel. Once a higher order unit is identified, the processing of component letters ceases.

Tests of the unitization model typically involve variation of conditions that will facilitate or inhibit large-unit

identification. The prediction is that when large unit identification is facilitated, then component-letter identification will be disrupted. Healy, Oliver, and McNamara (1987) found the expected pattern in a study in which subjects searched for the letter *t* in the word *the*, as well as in other words when the words containing the target were spelled correctly or incorrectly. More misses occurred when the target appeared in *the*, but only when the words were spelled correctly. That is, a high-frequency word will lead to fast word-unit identification, and the component-letter identification process will be truncated. Misspellings, however, will slow down the word-unit identification. Other race models make similar assumptions (e.g., Johnson, Allen, & Strand, 1989).

Recently, Hadley and Healy (1991) have shown that the disruptive effects of word unitization on component-letter identification are limited to parafoveal processing. Lexical processing, however, depends largely on foveal processing (see Rayner & Balota, 1989, for a review of foveal and parafoveal interactions); moreover, parafoveal effects seem to be mediated chiefly by abstract letter information (p. 279). Studies of foveal word identification tend to show facilitative effects of a word context on letter identification. For example, Reicher (1969) reported that forced-choice letter accuracy is better following a word display than it is following a display of a single letter or a string of letters forming a nonword. These facilitative effects of words on letter identification have triggered a variety of explanations. Connectionist models (e.g., McClelland & Rumelhart, 1981) assume that letter units receive activation from word-level units that single letters and letters in nonwords would not receive. Symbolic models (Richman & Simon, 1989) assume that "words are identified through letters . . . The only thing that gets reported to short-term memory is the chunk found" (p. 419). It would be premature to accept any single explanation of the Reicher-type phenomena, given that the

This research was supported by the Natural Sciences and Engineering Research Council of Canada (Grant 0684). The data were collected by Paul Greveling and Janice Caulfield; programming assistance was provided by Wray Hutton. Comments by Lester Krueger, Bart Farrell, and two anonymous reviewers on an earlier draft were helpful in strengthening the paper. Requests for reprints should be addressed to H. H. C. Marmurek, Department of Psychology, University of Guelph, Guelph, Ontario N1G 2W1, Canada (e-mail: psymarm@vm.uoguelph.ca).

context effects on letter processing are limited by methodological variations (Marchetti & Mewhort, 1986).

Whereas the accounts outlined above admit to some regularity in letter and word processing, there has been recent interest in episodic effects on word processing. There are a host of studies showing that prior experiences can facilitate the perceptual processing of words (see Tulving & Schacter, 1990, for a review of mechanisms underlying priming). The present experiments examined the conclusion of Whittlesea and Brooks (1988) that episodic experiences can "reverse the general advantage of one linguistic unit over another" (p. 397).

The episodic framework argues that the nature of a prior experience with a word will determine the units of perception on a later trial. The paradigm used to test the episodic framework involves presenting subjects with a training task and then a transfer test. For example, in their Experiment 6, Whittlesea and Brooks (1988) showed words on a monitor. In the training phase, subjects wrote down either entire words or a letter within the words (the letter was designated by a caret above the letter). In the transfer test, words were displayed briefly (30 msec). Prior to each test trial, subjects were told to identify either the entire word or a letter indicated by a caret in the test position.

The critical finding was a significant training \times test interaction. Whittlesea and Brooks (1988) described the interaction as follows: "reinstating the previously experienced processing context facilitates performance" (p. 395). The means entering the interaction were as follows: for word identification, word training yielded .80 accuracy and letter-in-word training yielded .60 accuracy; for letter-in-word identification, word training yielded .65 accuracy and letter-in-word training yielded .70 accuracy. No simple main effects tests were reported, but the *SE* for all means was about .04, which suggests that the word-identification means were significantly different but that the letter-in-word means were not. Clearly, the training had a weaker effect on the identification of letters in words than on identification of the whole word.

The task used in the transfer phase of Experiment 6, however, promoted word-unit activation, and lexical encoding may constrain the priming effects (Weldon, 1991). For example, when pseudowords (pronounceable and orthographically legal nonwords) served as the stimuli (Experiments 1-4), Whittlesea and Brooks (1988) found no advantage for reporting letters presented in pseudowords over letters presented alone (Experiment 1). In Experiment 4, training on single letters led to more accurate performance on single-letter identification than on the identification of letters in pseudowords. In contrast, single-letter identification was never better than letter-in-word identification in Experiment 6. Although the pseudoword-word comparisons are based on different experiments, they are consistent with the view that the transfer task tapped lexical processing when words were used as stimuli.

The specific issue addressed in the present study was whether or not training effects could be demonstrated for tasks in which lexical encoding was limited. One paradigm that apparently precludes word-unit activation is the simultaneous-matching task in which only the first letters of displays are to be compared. Two effects suggest that task limits word activation. First, letter-target decisions are made more quickly than are whole-string decisions. Secondly, first letters of words are compared no more quickly than are first letters of nonwords (Marmurek, 1989). Those outcomes are consistent with word-perception theories that assume attention may be directed to specific locations within a word prior to encoding the entire word (Johnson et al., 1989; Marchetti & Mewhort, 1986).

EXPERIMENT 1

Experiment 1 tested the episodic framework of word and letter perception by giving subjects a training phase similar to that used by Whittlesea and Brooks (1988). In the transfer task, subjects completed a *same-different* simultaneous-comparison task in which they judged the identity of either entire displays or the first letters in the displays. Both word and pseudoword displays were used. The transfer task also included new items, which had not been used during training, to serve as a baseline for learning effects due to any prior processing. According to the episodic framework, entire comparisons should be faster if training had involved writing down the entire display than if training had involved writing the first letter, and the reverse should occur for first-letter comparisons. Alternatively, if episodic effects are limited to conditions in which word units are activated, then first-letter comparisons should not show episodic effects.

Method

Subjects. The subjects were 96 introductory psychology students who participated in partial fulfillment of their course requirements. The subjects were assigned in an alternating order to the word and pseudoword display conditions. All subjects were native English speakers and had normal or correct-to-normal vision.

Design and Stimuli. Display (words vs. pseudowords) was varied between subjects. Training (whole, first, none) and test (whole vs. first) were varied within subjects. The stimuli were selected from those used by Marmurek (1986) and some additional items for *different* trials. The one-syllable five-letter words had a mean frequency of 108.09 (*SD* = 95.79) per million (Kučera & Francis, 1967); the pseudowords were generated from the words by replacing vowels so that the stimuli were regular in pronunciation (e.g., BRUSH was changed to BRISH).

Separate lists were constructed for subjects tested with either words or pseudowords, but the structure of the lists was the same for the two types of displays. There were 72 items in the study phase and 128 pairs of items in the transfer test. All displays shown in the study phase appeared on the transfer test. The transfer test also included items not shown during the study phase. Six transfer test lists were constructed in order to rotate items across the six training \times test conditions. Of the 128 pairs in a list, there were 64 items assigned to the *same* response and 64 items assigned to the *different* response. Among the 64 *same* pairs were 48 identical pairs.

Those pairs in a particular list were selected from six sets of eight words each, such that the sets were counterbalanced across lists to test the six training-test combinations. Eight subjects in each display group received each list.

Of the 48 identical pairs in a list, 24 were used on whole-comparison trials and 24 were used on first-letter trials. The 24 items on each type of trial included 8 items studied as whole items, 8 items studied as first-letter items, and 8 items that had not been studied. To ensure that first-letter trials were based on first-letter processing, an additional 16 pairs were used on first-letter trials where only the first letters were identical (e.g., COAST-CRIME). These 16 pairs consisted of two sets of 8 pairs, so that, in three lists, one set was presented during training and, in the other three lists, the other set of 8 was presented during training. Accordingly, of the 72 study items, 40 later appeared on *same* trials, 32 from the rotated sets (16 tested for whole comparisons and 16 tested for first-letter comparisons) and 8 from the first-only items. Analyses were restricted to the rotated sets of identical items. That is, the first-letter-only items were included to vary the criteria used on first-letter and whole-display trials.

The 64 *different* items were constructed so that decisions would be made on the basis of separate criteria for first-letter and whole-display trials. For first-letter decisions, displays differed in the first letter, and the identity of letters in other positions varied (e.g., BRAVE-GRAVE, MONTH-NORTH, and TRAIN-KNIFE). The number of instances of those variations was not varied systematically, but the inconsistency in the identity of noninitial letters, in conjunction with the *same* first-letter trials where only the first letter displays were identical, was designed to encourage subjects to attend to the first position. For *different* whole-comparison trials, the pairs differed in various letter positions including the first (e.g., BEACH-TEACH), the third (e.g., TRACK-TRICK), the fifth (GRAND-GRANT), the first and third (e.g., CHUNK-THINK), the first and fifth (e.g., CHEEK-SHEET), and all letters (e.g., WHEAT-SMOKE). The number of instances of these variations ranged from 2 to 5; they were intended to encourage subjects to examine all positions on whole-comparison trials. Marmurek (1986) showed that noninitial position information influenced whole comparisons but not first-letter comparisons. For the 32 pairs that included only one item that had been studied, the *old* item was equally likely to be the top or bottom item.

Practice study lists of 8 words or 8 pseudowords and practice test lists of 16 words or pseudowords were constructed for use prior to the experimental study and test lists. The practice test list included four nonstudied items and two instances of each of the six test-training combinations used in the experimental trials. Half of the practice items had *same* as the correct response.

Procedure. All displays were shown in white, centered on a black background of a Data Train DC-518 color monitor connected to a 286 PC-compatible microcomputer. The computer was programmed to present displays and record responses and latencies to the nearest millisecond. The displays were shown in uppercase at horizontal and vertical visual angles of 0.96° and 0.38°, respectively. Each letter was approximately 4 mm high and 2 mm wide. The subjects were seated about 60 cm from the screen.

In the training phase (both practice and experimental), a row of five asterisks was shown on the monitor for 0.5 sec and was followed by a blank screen for 0.5 sec. A cue was then presented for 0.5 sec. The cue was either the number 5, indicating a whole-display trial, or the number 1, indicating a first-letter trial. The cue offset was followed by a blank screen for 0.5 sec. Next, a display item (word or pseudoword) was shown for 0.5 sec and was followed by a blank screen for 3 sec. During that interval, the subjects were to write down either the entire display (if preceded by a 5) or the first letter of the display (if preceded by a 1). The subjects wrote the items on a lined sheet of paper numbered from 1 to 72. The end of the 3-sec response period was signaled by a brief tone, which was followed by the onset of the row of asterisks beginning the next trial.

Subsequent to the training phase, the subjects were reminded of the test-phase instructions given during the practice session. In the transfer test, the subjects were required to make whole-display and first-letter decisions. On each trial, a row of five asterisks appeared on the screen for 0.5 sec, followed by a blank screen for 0.5 sec. Then, a cue to indicate either a whole-display (5) or first-letter (1) decision was shown for 0.5 sec, after which the screen was blank for 0.5 sec. Next, the comparison items were shown simultaneously, one above the other with a vertical separation of 4 mm (vertical visual angle = 1.15°), until the subject responded. Responses were made by pressing one of two microswitches marked S for *same* and D for *different*. The microswitches were placed 10 cm apart and were separated by a resting button on which the subjects placed the forefinger of their preferred hand.

Results and Discussion

The mean error rates (in %) and the mean correct response times (RTs; in milliseconds) for *same* responses and for *different* responses are shown in Table 1 and Table 2, respectively. The latencies were trimmed by excluding, for each subject, the trials on which a response exceeded the mean for a condition by more than 2 *SD*. The positive correlation between error rate and mean correct RT across the combination of conditions for words [$r(10) = .68$] and pseudowords [$r(10) = .66$] suggests there was no speed-accuracy tradeoff. There was no evidence that the position of the *old* (top or bottom) influenced the pattern of effects. A 2 (display) \times 3 (training) \times 2 (test) mixed analysis of variance (ANOVA), with display as the only between-subject factor, was conducted

Table 1
Mean Correct *Same* Latencies and
Error Rates in Experiment 1

Comparison	Training					
	First Letter		Whole		None	
	RT	ER	RT	ER	RT	ER
Word Display						
First Letter	777	5.22	768	3.16	783	3.93
Whole display	918	4.69	921	3.65	956	4.82
Pseudoword Display						
First Letter	752	4.44	757	3.52	759	5.23
Whole Display	1,007	5.21	1,018	6.25	1,025	5.81

Note—RT = mean correct *same* latency in milliseconds. ER = percent error rate.

Table 2
Mean Correct *Different* Latencies and
Error Rates in Experiment 1

Comparison	Training					
	First Letter		Whole		None	
	RT	ER	RT	ER	RT	ER
Word Display						
First Letter	853	2.60	832	2.08	859	3.37
Whole display	932	4.69	1,033	10.42	942	5.98
Pseudoword Display						
First Letter	837	3.25	826	2.08	805	2.33
Whole Display	893	4.87	987	7.85	925	6.49

Note—RT = mean correct *different* latency in milliseconds. ER = percent error rate.

on the mean correct RTs and the mean percent errors. Although a full understanding of the processes contributing to *same* and *different* judgments remains controversial (Farell, 1985), it is generally found that task variables interact with the two judgments, suggesting that separate processes govern those judgments (Pellegrino, Doane, Fischer, & Alderton, 1991). Accordingly, separate analyses were conducted for each type of judgment.

Same trials: Error rates. The analysis of error rates showed that significantly more errors were made for whole comparisons than for first-letter comparisons, [$F(1,46) = 6.59, MS_e = 50.18, p < .05$]; however, the display \times test interaction was also significant [$F(1,94) = 4.31, p < .05$]. For words, whole comparisons (M error rate = 4.39) led to 0.29% more errors than did first-letter comparisons (M error rate = 4.10); for pseudowords, the difference was 1.36% (M error rates for whole-display and first-letter comparisons were 5.76 and 4.40, respectively). The lower error rate for whole-trained items ($M = 4.15$) relative to letter-trained items ($M = 4.89$) and new items ($M = 4.95$) was not significant [$F(2,188) = 2.07, MS_e = 38.05, p > .10$], and the test \times training interaction was not significant [$F(2,188) = 1.68, MS_e = 38.70, p > .15$].

Same trials: Mean correct latencies. A 2 (display) \times 3 (training) \times 2 (test) analysis showed a significant effect of test [$F(1,94) = 78.49, MS_e = 28,402.07, p < .001$]. Judgments were faster for first-letter comparisons ($M = 766$ msec) than for whole-display comparisons ($M = 974$ msec). The test effect, however, interacted with display [$F(1,94) = 9.14, p < .01$]. For words, the whole-display comparisons ($M = 932$ msec) were 156 msec longer than first-letter comparisons ($M = 776$ msec). For pseudowords, the whole-display comparisons ($M = 1,017$ msec) were 262 msec longer than first-letter comparisons ($M = 756$ msec). The interaction also showed that the simultaneous-matching task was facilitated by lexicality on whole-display trials ($M = 932$ msec, for words; $M = 1,017$ msec, for pseudowords), but not on first-letter trials, where the pattern was reversed ($M = 776$ msec, for words; $M = 756$ msec, for pseudowords). The absence of the familiarity benefit on first-letter comparisons suggests that lexical encoding is not preliminary to letter activation when attention is directed to a pre-designated position within a word (Marmurek, 1989).

The overall training effect on *same* trials was weak [$F(2,188) = 2.73, MS_e = 8,595.73, p < .07$]. The judgments for new items ($M = 881$ msec) were longer than those for old items ($M = 865$ msec). The advantage of old items depended on the type of test [$F(2,188) = 3.65, p < .05$]. For whole comparisons, the advantage was 25 msec (991 vs. 966); for first-letter comparisons, the advantage was 7 msec (771 vs. 764). The three-way interaction of display \times test \times training was marginally significant [$F(2,188) = 2.61, p < .10$].

To explore that trend, separate 2 (test) \times 3 (training) analyses were run for each type of display. For words, the effect of test was significant [$F(1,47) = 32.35, MS_e =$

24,742.28, $p < .001$], as was the effect of training [$F(2,94) = 3.18, MS_e = 6,398.27, p < .05$]. The interaction of test \times training [$F(2,94) = 2.89, MS_e = 7,719.65, p < .10$] suggests that the benefit of prior study was greater on whole-word comparisons (M advantage for the pooled old items vs. the new items = 36 msec) than on first-letter comparisons (M advantage = 10 msec). The analysis of pseudoword latencies yielded only an effect of test [$F(1,47) = 64.40, MS_e = 40,223.70, p < .001$]. First-letter comparisons were faster than whole-display comparisons. For the main effect of training and for the test \times training interaction, $F(2,94) < 1$. The distinct training patterns for words and pseudowords indicate that training effects were mediated by lexical encoding. The test \times training interaction for words is consistent with the hypothesis that lexical encoding influences whole-display comparisons to a greater extent than it influences component-letter comparisons (Marmurek, 1989). Note, however, that there was no indication that compatibility between training and test facilitated the whole-word judgments, in that first-letter and whole-word training yielded similar effects.

Different trials: Error rates. A 2 (display) \times 2 (test) \times 3 (training) analysis showed that the main effect of test was significant [$F(1,94) = 16.42, MS_e = 57.77, p < .001$]. The error rate was higher on whole-display comparisons ($M = 6.72\%$) than on first-letter comparisons ($M = 2.62\%$). The test \times training interaction was significant [$F(2,188) = 13.67, MS_e = 51.11, p < .001$]. Relative to new items on whole-comparison trials ($M = 6.24\%$), old items that had been written as whole items in training led to *more* errors ($M = 9.14\%$); old items for which the first letter had been written led to fewer errors ($M = 4.78\%$). Relative to new items on first-letter comparisons ($M = 2.85\%$), however, there was a small reduction for items written as wholes ($M = 2.08\%$) and a small increase for items whose first letters had been written ($M = 2.93\%$).

The three-way interaction of display \times test \times training was also significant [$F(2,188) = 3.68, p < .05$]. To explore the training effects for each type of display, separate 2 (test) \times 3 (training) analyses were run for the word and pseudoword groups. For the word group, the test \times training interaction was significant [$F(2,94) = 3.93, MS_e = 36.71, p < .05$]. Contrasts between the new items and the old items showed that, for whole comparisons, the items trained as whole words led to significantly more errors [$F(1,94) = 12.87, p < .001$]. There were no other significant contrasts with the new item within a given type of test. For the pseudoword group, the test \times training interaction was also significant [$F(2,94) = 5.41, MS_e = 78.55, p < .01$]. However, unlike the differences among words, no combination of test and training yielded a significant difference from new pseudowords. Whole comparisons for pseudowords showed a pattern similar to that for words, in that items trained as wholes led to marginally *more* errors than did items for which first letters had been written down [$F(1,94) = 2.76, p < .10$]. The most

direct way to describe the three-way interaction, then, is that whereas first-letter comparisons were unaffected by prior training, whole comparisons were impeded by prior whole training, particularly for words. On whole-display comparisons, it is possible that lexical encoding during training led the subjects to recognize one of the items in a *different* pair. That recognition might have led the subjects to prepare for a *same* response. The source of the interference then might have been either response interference or a higher criterion for rechecking disconfirming information between the items in a pair (Krueger, 1978).

Different trials: Mean correct latencies. A 2 (display) \times 2 (test) \times 3 (training) analysis showed that latencies were faster for first-letter decisions ($M = 835$ msec) than for whole-display decisions ($M = 952$ msec) [$F(1,94) = 58.35$, $MS_e = 12,987.58$, $p < .001$]. The test \times training interaction was significant [$F(2,188) = 18.39$, $MS_e = 10,208.40$, $p < .001$], as was the display \times test \times training interaction [$F(2,188) = 4.42$, $p < .02$]. Separate analyses for words and pseudowords showed that the test \times training interaction was significant for both [for words, $F(2,94) = 3.59$, $MS_e = 17,383.46$, $p < .05$; for pseudowords, $F(2,94) = 9.19$, $MS_e = 6,392.29$, $p < .001$]. Contrasts with new items showed that, for words, whole comparisons were slower for old words trained as whole words (M increase = 91 msec) [$F(1,94) = 14.08$, $p < .001$]. No other contrast against new items was significant within a given type of test. Similarly, for pseudowords, the only significant contrast against new items was for old items trained as wholes in the whole-display test condition (M increase = 62 msec) [$F(1,94) = 9.04$, $p < .001$]. The increases in the first-letter comparisons were not significant ($p > .10$).

The increase in latency to whole-display decisions about words trained as words complemented the increase in error rates in those conditions. If whole training led to lexical encoding, then recognition of an item represented by the encoded lexical unit on the transfer test might bias subjects to prepare a *same* response. Consequently, response competition might slow down the selection of a *different* response or lead to an incorrect response. Alternatively, the criterion for a *different* response would be raised so that either more time would be taken to recheck the items on the test (leading to a longer latency) or the criterion would not be met (leading to more errors). These effects would not occur on first-letter tests, because rechecking is limited to a single position where the criterion may not be as variable, nor would they occur with first-letter training, which limits the original encoding of the lexical units. The items would be less likely to be recognized on the transfer test following first-letter training (Marmurek, 1992). The similar, but weaker, effects with pseudowords suggest that pseudowords were also remembered better after whole-display and first-letter training. Because the pseudowords were orthographically regular and pronounceable, it may be that the pseudoword

effects were mediated by some of the mechanisms that underlie lexical priming (Rueckl, 1990, p. 387).

EXPERIMENT 2

The critical findings of Experiment 1 were that lexicality and training affected whole-display comparisons, but not first-letter comparisons. Overall, first-letter decisions were made more quickly than were whole-display decisions. These results suggest that that comparison task tapped letter-level processing; lexical encoding was more likely for whole-display comparisons, and only under the latter condition was training important. Consistent with Whittlesea and Brooks (1988), whole-word comparisons were more sensitive to whole-word training than to single-component-letter training. However, in contrast to Whittlesea and Brooks (1988), there was no evidence that component-letter training facilitated letter processing. Recall that the training effect for letter identification reported by Whittlesea and Brooks was small and equivalent to the standard error of the mean.

The purpose of Experiment 1 was to show that the resistance of letter-in-word processing to facilitation was not specific to a task that promoted whole-word processing. To the extent that first-letter and whole-display tasks were differentially sensitive to lexicality, those tasks promoted either sublexical or lexical processing. Only in the lexical conditions was there evidence of episodic effects. The chief purpose of Experiment 2 was to replicate the finding that episodic effects were specific to lexical processing. The inclusion of pseudowords in Experiment 1 had been designed chiefly to reveal the role of lexical processing in the whole-display task; moreover, because the training effects were larger for words than for pseudowords, only words were used in the second experiment.

The major procedural change was to vary training between groups of subjects. There were three reasons for that change. First, a pure list-training manipulation might encourage whole-word subjects to rely on memory during the transfer test; in the mixed-list design of Experiment 1 that strategy might have been avoided in the face of poor memory for first-letter study items. Second, the pure list-training manipulation might facilitate first-letter subjects in attending to the component letter during training. More first-letter training might then lead to facilitation of first-letter decisions in the transfer test. Finally, the pure training list made it easier to counterbalance items across conditions so that more items could be tested in each of the critical conditions.

Method

Subjects. The subjects were 48 undergraduate students in an introductory psychology course who participated in partial fulfillment of a course requirement. The students were assigned to the whole-word and first-letter training groups in alternating order. All subjects were native English speakers and had normal or corrected-to-normal vision.

Design and Materials. Training (first letter vs. whole word) was varied between subjects. Pairs of subjects were yoked to transfer lists in which the items formed conditions defined by the combination of test (first letter vs. whole word), episode (old vs. new), and response (*same* vs. *different*). The items were selected from those used in Experiment 1. The study list consisted of 64 words, and the transfer test list consisted of 128 word pairs divided into 64 *same* pairs and 64 *different* pairs.

The *same* pairs included 48 pairs that contained identical words and 16 pairs that were identical only in their first letters. The identical-word pairs were rotated across subjects to appear equally often on first-letter and whole-display trials. The 16 *same* pairs identical in only their first letters occurred on first-letter trials. The 64 *same* pairs were divided into two sets, each comprising 24 identical pairs and 8 first-only pairs; across subjects, each set appeared equally often on the study trials. For the first-only pairs, the subjects studied only one of the items. None of the items presented on the *new* trials had been studied. Because training (first letter or whole word) was varied between subjects, there were 16 *same* items per test \times episode condition, in contrast to Experiment 1, in which there were 8 *same* items per condition.

The 64 *different* trials on the transfer test included 32 items presented during the study phase. There were 16 *different* items per training \times test \times episode condition. The differences between the items in a pair followed the same constraints as in Experiment 1 for first-letter and whole-word trials. Separate lists ensured that each item was tested equally often as an old or new item.

Procedure. The between-group manipulation in the training phase made the cue during study redundant. Therefore, the 1 and the 5 cues that appeared during training in Experiment 1 were deleted in Experiment 2. Prior to the training list, the subjects were told to write down either the entire word or the first letter of each of the words they would see. In all other respects, the procedure used in Experiment 2 was the same as that used in Experiment 1.

Results and Discussion

The mean correct decision latencies (in milliseconds) and the mean error rates (in %) for *same* and *different* responses are shown in Tables 3 and 4, respectively. The positive correlation between the two measures across all combinations of conditions [$r(14) = .49$] indicates that there was no speed-accuracy tradeoff. The dependent measures were submitted to 2 (training, first letter vs. whole word) \times 2 (test, first letter vs. whole word) \times 2 (episode, old vs. new) mixed factorial analyses in which training was the only between-subject variable. As in Experiment 1, separate analyses were run for the *same* and *different* trials.

Table 3
Mean Correct *Same* Latencies and Error Rates in Experiment 2

Episode	Training			
	First Letter		Whole Word	
	RT	ER	RT	ER
First-Letter Comparison				
Old	750	5.94	782	7.50
New	764	5.88	794	6.38
Whole-Word Comparison				
Old	826	5.13	840	2.63
New	839	5.06	920	8.25

Note—RT = mean correct *same* latency in milliseconds. ER = percent error rate.

Table 4
Mean Correct *Different* Latencies and Error Rates in Experiment 2

Episode	Training			
	First Letter		Whole Word	
	RT	ER	RT	ER
First-Letter Comparison				
Old	773	1.88	793	3.38
New	768	4.31	798	4.34
Whole-Word Comparison				
Old	857	6.99	941	7.63
New	846	6.75	876	6.19

Note—RT = mean correct *different* latency in milliseconds. ER = percent error rate.

Same trials: Error rates. The overall mean error rate was 5.85%. The ANOVA showed no significant main effects, but a marginally significant interaction of training \times test \times episode [$F(1,46) = 3.99$, $MS_e = 22.84$, $p < .06$]. Inspection of the means in Table 3 suggests that the fewest errors occurred for old items tested as whole words following whole-word training ($M = 2.63$). To test this interpretation, separate analyses were run for the two groups of subjects. For the group that received first-letter training, the main effects and interaction were not significant [all $F_s(1,23) < 1$]. For the group that received whole-word training, the episode effect was significant [$F(1,23) = 5.87$, $p < .05$], as was the interaction of test \times episode [$F(1,23) = 8.49$, $p < .01$]. For whole-word comparisons, old items led to a reduction of 5.62 in percentage errors relative to new items [$F(1,23) = 17.71$, $p < .001$]; for first-letter comparisons, there was a small and unreliable difference in the error rates for old and new items [$F(1,23) < 1$]. The reduction in errors to old items trained as wholes on the whole-comparison task suggests that *same* responses were primed by the recognition of those items (Eriksen, O'Hara, & Eriksen, 1982). This priming may have offset any internal noise that normally interferes with a *same* response (Krueger, 1978). Whatever the source of the episodic effects, they did not occur following first-letter training, because that training may be less likely to lead to lexical encoding; the episodic effects also did not occur on first-letter comparisons following whole-word training, because first-letter comparisons truncate whole-word processing on the transfer test.

Same trials: Mean correct latencies. A 2 (training) \times 2 (comparison) \times 2 (episode) mixed ANOVA of the mean correct latencies shown in Table 3 revealed that the two within-subject main effects were significant [for test, $F(1,46) = 39.37$, $MS_e = 5,643.77$, $p < .001$; for episode, $F(1,46) = 12.75$, $MS_e = 2,233.81$, $p < .001$]. These effects showed that decisions were made more quickly for first-letter comparisons ($M = 773$ msec) than for whole-word comparisons ($M = 856$ msec) and that decisions were faster for old words ($M = 800$ msec) than for new words ($M = 829$ msec). However, both effects entered into the significant three-way interaction of train-

ing \times test \times episode [$F(1,46) = 4.46$, $MS_e = 2,238.76$, $p < .05$]. Inspection of Table 3 suggests that the episodic effects were largest for the whole-word training group on whole-word comparisons. Separate analyses for each group confirmed this interpretation. For the first-letter training group, the main effect of episode was not statistically significant [$F(1,23) = 2.81$, $p > .10$]. The main effect of test was significant [$F(1,23) = 19.42$, $p < .001$], but the test \times episode interaction was not [$F(1,23) < 1$]. For the whole-word training group, all three effects were significant [for episode, $F(1,23) = 9.94$, $p < .01$; for test, $F(1,23) = 20.21$, $p < .001$; for the text \times episode interaction, $F(1,23) = 5.69$, $p < .05$]. Decisions were faster for the old items than for the new items on whole-word comparisons [$F(1,23) = 18.37$, $p < .001$], but not on first-letter comparisons [$F(1,23) < 1$].

The delay in judging new items as the *same* on whole-word comparisons is consistent with the hypothesis that episodic effects are lexical. First-letter training yields less reliance on prior lexical encoding than does first-letter testing. On whole-word comparisons following whole-word training, subjects may use their memory of the study words to guide their decisions. Recognition of an item as old may prime the *same* response, which is compatible with the displays. New items may have primed a *different* (or *no*) response that would lead to further rechecking of the displays.

Different trials: Error rates. The mean overall error rate on *different* trials was 5.18%. The 2 (training) \times 2 (test) \times 2 (episode) analysis showed that only the effect of test was significant [$F(1,46) = 14.99$, $MS_e = 24.99$, $p < .001$]. Fewer errors occurred on first-letter comparisons ($M = 3.48\%$) than on whole-word comparisons ($M = 6.89\%$). The only other effect to yield $F(1,46) > 1$ was the test \times episode interaction [$F(1,46) = 2.47$, $.10 < p < .15$]. The trend indicated by this interaction, although not statistically significant, was that the relative difficulty of whole-word comparisons over first-letter comparisons increased for old items. This pattern complements the lower error rates on *same* trials and reinforces the hypothesis that recognition of a word might have primed a *same* response that conflicted with the correct *different* response. By raising the criterion for rechecking, differences were missed. Note, however, that, for the *different* trials, the original training did not moderate that effect. The absence of the significant interactions on the *different* trials is consistent with the proposal of separate processes that govern *same* and *different* decisions (Farell, 1985).

Different trials: Mean correct latencies. A 2 (training) \times 2 (test) \times 2 (episode) mixed analysis of the means shown in Table 4 yielded significant main effects for test [$F(1,46) = 51.69$, $MS_e = 5,777.56$, $p < .001$] and for episode [$F(1,46) = 6.33$, $MS_e = 1,821.88$, $p < .05$]. These effects showed that latencies were faster for first-letter comparisons ($M = 783$ msec) than for whole-word comparisons ($M = 880$ msec), and were faster for new items ($M = 822$ msec) than for old items ($M = 841$ msec).

Note that the episodic effect here was one of interference. The test \times episode interaction was significant [$F(1,46) = 4.19$, $MS_e = 2,718.49$, $p < .05$], and it showed that the episodic effect occurred for whole-word comparisons (M difference = 38 msec) but not for first-letter tests (M difference = 0 msec). The three-way interaction of training \times test \times episode was marginally significant [$F(1,46) = 3.98$, $p < .10$].

Because the effects of training were central to the predictions, the interaction was probed by running separate analyses for each training group. For the first-letter group, only the main effect of test was significant [$F(1,23) = 13.11$, $p < .01$]. For the interaction, $F(1,23) < 1$. The whole-word training group, however, showed a significant effect of test [$F(1,23) = 55.59$, $p < .001$], a significant effect of episode [$F(1,23) = 4.84$, $p < .05$], and a significant interaction [$F(1,23) = 9.37$, $p < .01$]. The difference between old and new items was not significant for first-letter comparisons [$F(1,23) < 1$]; the 65-msec delay for old items, relative to new items, was significant for whole word comparisons [$F(1,23) = 16.69$, $p < .001$]. Training whole words slowed down *different* comparisons of old items on whole-word comparisons. There were no comparable episodic effects on first-letter comparisons.

GENERAL DISCUSSION

The major finding of the two experiments was that whereas decisions about the identity of words were influenced by a single prior experience with those words, decisions about the identity of a single letter within the words were insensitive to episodic effects. This general pattern held for both error rates and latencies, and for *same* decisions as well as for *different* decisions.

Although the critical effects were not statistically significant in each experiment, the direction of the effects was consistent across experiments. For example, in Experiment 1, the 1.17 difference in percentage error between whole-trained and new items did not enter into a significant test \times training interaction; however, the interaction including the 5.62 difference was significant in Experiment 2. Combining the effects of the two experiments (Rosenthal & Rosnow, 1991, p. 497) yielded $Z = 2.77$, $p < .01$ (two-tailed). Conversely, Experiment 1 showed that significantly more errors were made on old words, relative to new words, on *different* trials when the old words had received whole-word training (M difference = 4.44) rather than first-letter training (M difference = -1.29); in Experiment 2, this interaction was not significant (the respective mean differences were 1.44 and 0.24). However, combining the interaction effects across the experiments yielded $Z = 2.49$, $p < .02$. Thus, the pattern of results was robust across two experiments using different designs (training was manipulated within subjects in Experiment 1 and between subjects in Experiment 2).

The two types of responses showed complementary effects. For *same* responses, the episodic effects were facilitative; for *different* responses, the episodic effects were inhibitory. The opposing patterns on whole-comparison trials for the two types of responses suggest that recognition of an old item in the pair of comparison items might have influenced distinct processes that govern the two types of responses. In attempting to encode both items, recognition of an item as *old* may have biased the subjects to prepare a *same* response. The processing effect of this bias would be to limit the number of letters that are rechecked prior to a response (Krueger, 1984, p. 283). Focused rechecking of familiar items would then yield relatively fast and accurate decisions on *same* trials. For *different* pairs that included an old item, however, focused rechecking would lead to an error if only letters common to the two items were rechecked. The relatively long latencies for correct old-*different* trials may be due to the extra processing necessitated by the conflict of a primed *same* response following recognition of one of the items and the detection of some difference. Given that conflict, subjects might carry out a more exhaustive rechecking of the displays.

The limitation of episodic effects to whole-display comparisons contradicts the claim that letter identification depends on a particular prior experience with that letter in a word (Whittlesea & Brooks, 1988). In the Whittlesea and Brooks tachistoscopic identification task, whole-word identification generally was superior to letter-in-word identification; moreover, they found that episodic facilitation was greater for whole-word reports (5 *SE*) than for letter-in-word reports (1 *SE*). The rationale for using the speeded simultaneous-matching task in the present experiments was to determine whether letter-in-word facilitation could be enhanced in a task in which letter-in-word performance was superior to whole-word performance. That is, the weak letter-in-word facilitation found by Whittlesea and Brooks might have been due to subjects' learning to attend to a particular position within a word during training so that, on the transfer test, the episodic effect on the letter *followed* recognition of the word. The present results suggest that letter-in-word performance is resistant to episodic effects in the absence of prior lexical encoding.

The theoretical importance of these results relates to the issue of the primary units of word perception. Whittlesea and Brooks (1988) accepted the notion, advanced from studies of speech perception, that no linguistic unit enjoys universal psychological primacy (McNeill & Lindig, 1973); rather, in a given task, attention will be allocated to particular units. McNeill and Lindig showed that the time taken to detect a phoneme, syllable, or word depended on the linguistic context of the comparison items so that detection was faster when the level of the target and the comparison items matched (e.g., phonemes were detected more quickly in a list of phonemes than in a list of syllables). Norris and Cutler (1988) argued that the phoneme conditions studied by McNeill and Lindig were

really syllable conditions (the phoneme targets were initial consonants preceding the vowel /a/) and showed that target-to-foil similarity determined the relative speed of phoneme and syllable monitoring. Note, however, that McNeill and Lindig (1973) found that syllables were detected faster in a list of words than in a list of syllables (Table 2, p. 422, and Table 3, p. 423). Therefore, the matching of target linguistic level and comparison item linguistic level does not provide a complete account for their results. Also, over all levels of matches, the fastest detections were of phonemes. The McNeill and Lindig results, then, are consistent with the present findings in suggesting that attention may be allocated to sublexical units prior to the encoding of the lexical unit.

Whittlesea and Brooks (1988) concluded that the perception of any linguistic unit could be influenced by prior experiences with that unit. They contrasted general versus specific models of word processing and faulted the former's inability to account for the flexibility due to prior specific experiences. However, priming effects have triggered a variety of theoretical accounts that may be classified as connectionist or episodic (see Rueckl, 1990, for a review of the theories and their indeterminacy, p. 388). The accounts share the assumption that some interaction of visual, orthographic, phonological, and semantic information underlies the priming effects. That is, one of the strengths of these general structural models is to account for specific episodic effects.

The limitation placed on the models by the present results is that, when attention is allocated to a particular position in a display so that its lexical status is psychologically irrelevant, the complex interaction of various higher order sources of information that underlie lexical priming may be bypassed. That is, when letter-in-word identification is faster than whole-word identification, the mechanisms responsible for lexical priming will not be activated to an extent sufficient to yield episodic effects on the letters within words. This hypothesis is consistent with the view that the distribution of attention will determine whether letter units or holistic units achieve primacy in a given task (Marmurek, 1989; McNeill & Lindig, 1973; Whittlesea & Brooks, 1988). That attention is not directed at word units in the first-letter task was shown by the absence of a word advantage over pseudowords in this condition. Contrary to the position advocated by Whittlesea and Brooks, the allocation of attention to a letter within a word in the absence of lexical encoding is not a process that is sensitive to prior processing effects.

The lack of episodic effects on first-letter comparisons may be an artifact of a methodological difference between the status of old and new words and first letters. For words, *old* and *new* referred to the absence or presence of a test word in the study phase. For letter comparisons, *old* and *new* were not defined by the absence or presence of the letter during study but were defined by the absence or presence of the word containing the letter. That is, the methodological issue turns on the uniqueness of words relative to the repetition of letters across words, a dis-

tion due to the limited availability of different letters in the English alphabet. It is not clear whether priming of letters would emerge if only a subset of letters occurred as first letters during study. Any such priming would have to be examined with particular reference to the effect of using words with a limited set of first letters on whole-word comparisons. Moreover, the number of trials available per subject in a test of this issue would be small, given that studied and new letters would need to be tested on both *same* and *different* trials without repetition. The implications of this manipulation are not limited to the present study, however, in that the repetition of letters in the conditions formed by crossing the training and test variables was also characteristic of the Whittlesea and Brooks (1988) study. Nonetheless, the methodological issue does not mitigate the conclusion that a prior episode with a specific word leads to priming when that word is later processed as a whole word, rather than when attention is focused on only the first letter of the word.

It might be argued that requiring readers to attend to the first letter in a word lacks ecological validity. Attending to a whole word may be a more common tendency and one that facilitates identification of letters in words (as shown by word-superiority effects). Whole-word attention would lead to the detection of lexical violations, such as spelling errors (e.g., "seperate"), and, as shown in the present study, is subject to training effects. The utility of experiments that include tasks that demand attention to letters within words may lie in their potential to advance our understanding of attentional and letter-specific processing characteristic of early reading development and impaired reading (Riddoch, Humphreys, Cleton, & Fery, 1990). The results emerging from tasks that focus on letters within words also point to constraints on episodic effects in word perception. They will not occur when a letter enters consciousness separately from the context in which it is embedded.

REFERENCES

- ERIKSEN, C. W., O'HARA, W. P., & ERIKSEN, B. (1982). Response competition effects in *same-different* judgments. *Perception & Psychophysics*, *32*, 261-270.
- FARELL, B. (1985). "Same"- "Different" Judgments: A review of current controversies in perceptual comparisons. *Psychological Bulletin*, *98*, 419-456.
- HADLEY, J. A., & HEALY, A. F. (1991). When are reading units larger than the letter? Refinement of the unitization reading model. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *17*, 1062-1074.
- HEALY, A. F., OLIVER, C., & McNAMARA, T. P. (1987). Detecting letters in continuous text: Effects of display size. *Journal of Experimental Psychology: Human Perception & Performance*, *13*, 279-290.
- HUEY, E. B. (1908). *The psychology and pedagogy of reading*. Cambridge, MA: M.I.T. Press.
- JAMES, W. (1961). *Psychology: The briefer course*. New York: Harper & Row. (Original work published 1892)
- JOHNSON, N. F., ALLEN, P. A., & STRAND, T. L. (1989). On the role of word frequency in the detection of component letters. *Memory & Cognition*, *17*, 474-482.
- KRUEGER, L. E. (1978). A theory of perceptual matching. *Psychological Review*, *85*, 278-304.
- KRUEGER, L. E. (1984). Self-termination in *same-different* judgments: Multiletter comparison with simultaneous and sequential presentation. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *10*, 271-284.
- KUČERA, H., & FRANCIS, W. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.
- MARCHETTI, F. M., & MEWHORT, D. J. K. (1986). On the word-superiority effect. *Psychological Research*, *48*, 23-35.
- MARMUREK, H. H. C. (1986). Whole and part comparisons of words and nonwords. *Memory & Cognition*, *14*, 113-123.
- MARMUREK, H. H. C. (1989). Familiarity effects and word unitization in visual comparison tasks. *Memory & Cognition*, *17*, 483-489.
- MARMUREK, H. H. C. (1992, August). *Whole and partial word processing effects in implicit memory*. Paper presented at the meeting of the American Psychological Association, Washington, DC.
- MCCLELLAND, J. L., & RUMELHART, D. E. (1981). An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review*, *88*, 375-407.
- MCNEILL, D., & LINDIG, K. (1973). The perceptual reality of phonemes, syllables, words, and sentences. *Journal of Verbal Learning & Verbal Behavior*, *12*, 431-461.
- NORRIS, D., & CUTLER, A. (1988). The relative accessibility of phonemes and syllables. *Perception & Psychophysics*, *43*, 541-550.
- PELLEGRINO, J. W., DOANE, S. M., FISCHER, S. C., & ALDERTON, D. (1991). Stimulus complexity effects in visual comparisons: The effects of practice and learning context. *Journal of Experimental Psychology: Human Perception & Performance*, *17*, 781-791.
- RAYNER, K., & BALOTA, D. A. (1989). Parafoveal preview effects and lexical access during eye fixations in reading. In W. Marslen-Wilson (Ed.), *Lexical representation and process* (pp. 261-290). Cambridge, MA: MIT Press.
- REICHER, G. M. (1969). Perceptual recognition as a function of meaningfulness of stimulus material. *Journal of Experimental Psychology*, *81*, 274-280.
- RICHMAN, H. B., & SIMON, H. A. (1989). Context effects in letter perception: Comparison of two theories. *Psychological Review*, *96*, 417-432.
- RIDDOCH, J., HUMPHREYS, G., CLETON, P., & FERY, P. (1990). Interaction of attentional and lexical processes in neglect dyslexia. *Cognitive Neuropsychology*, *7*, 479-517.
- ROSENTHAL, R., & ROSNOW, R. L. (1991). *Essentials of behavioral research: Methods and data analysis* (2nd ed.). New York: McGraw-Hill.
- RUECKL, J. G. (1990). Similarity effects in word and pseudoword repetition. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *16*, 374-391.
- TULVING, E., & SCHACTER, D. L. (1990). Priming and human memory systems. *Science*, *247*, 301-306.
- WELDON, M. S. (1991). Mechanisms underlying priming on perceptual tests. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *17*, 526-541.
- WHITTLESEA, B. W. A., & BROOKS, L. R. (1988). Critical influence of particular experiences in the perception of letters, words, and phrases. *Memory & Cognition*, *16*, 387-399.

(Manuscript received February 21, 1992;
revision accepted for publication May 19, 1992.)