Psychological Review
1990, Vol. 97, No. 3, 447–452

More Words but Still No Lexicon: Reply to Besner et al. (1990)

Mark S. Seidenberg
McGill University
Montreal, Quebec, Canada

James L. McClelland
Carnegie Mellon University

The major points in the Besner, Twilley, McCann, and Seergobin (1990) critique of the Seidenberg and McClelland (1989) model are addressed. The model's performance differs from that of people in ways that are predictable from an understanding of the limitations of the implementation. The principal limitations are the size of the training corpus and the phonological representation. The issue of pseudohomophone effects is discussed, and Besner et al.'s new data are shown to be consistent with the Seidenberg and McClelland account of lexical decision.

Besner, Twilley, McCann, and Seergobin's (1990) broad critique of the Seidenberg and McClelland (1989) model merits closer inspection; we think the model stands up quite well. The model correctly simulates a broad range of behavioral phenomena; its performance departs from that of people in ways that are predictable from an understanding of limitations of the implementation. These limitations do not call into question any of the basic assumptions of the model.

Size of the Training Corpus

Besner et al. (1990) noted that the model does not perform as well as people on nonwords. They are correct. People know more than the model. The principal difference between the model and people is that whereas people's vocabularies are on the order of 30,000 words, the model's vocabulary is 2,897. This factor limits nonword performance. The model's performance on any given word is largely determined by the number of exposures to it during the training phase. There are also small effects due to neighbors sharing the same word body (e.g., GAVE/SAVE), as in the behavioral data we simulated. As the frequency of exposure to a word decreases, dependence on the neighbors increases. In the limit—a nonword on which the model has not been trained—performance is wholly determined by the neighbors. Hence, the limit on the size of the training corpus has a large effect on nonwords but very little effect on words. That is why our simulations focused on words.

We examined the model's performance on three sets of nonwords: those used by Glushko (1979, Experiment 2), those used by McCann and Besner (1987), and a set derived from the regular and exception words used in the Taraban and McClelland (1987) study (see Seidenberg & McClelland, 1989, Figures 12–14). Besner et al. (1990) noted that the model performs poorly on the Glushko nonwords. The problem items are listed in Table 1. The first point to note is that there is an ambiguity in how to score the data. Besner et al. scored responses such as KEAD = /ked/ as errors. However, in the training corpus, there are more neighbors in which -EAD is pronounced /ed/ than /Ed/. The model, then, was trained that /ked/ *is* the regular pronunciation, which it produced correctly. There are three such items in the Glushko list. For another item, GOMB, it is not clear what the regular pronunciation is. The four words in the -OMB neighborhood have three pronunciations (TOMB, BOMB, COMB, and WOMB); the model picked /Om/. Excluding these items, the model made 13/52 errors (25%).[1] In contrast, the model made 13/96 errors (13.5%) on the Taraban and McClelland (1987) nonwords and 66/160 errors (41.3%) on the McCann and Besner stimuli. Why performance differs so much across stimuli can be seen by examining how they relate to words in the training set. For each nonword, we counted the number of items in the corpus that have the same word body (e.g., for MAVE, all the -AVE words). The Taraban and McClelland stimuli have the most neighbors (these are items like MAVE and BINT), and the McCann and Besner stimuli have the fewest (these are items like VAWX and FAIJE). As Figure 1 indicates, the model's performance is related to the number of neighbors.

Our theory is that knowledge of spelling-sound correspondences derived from exposure to words is used in naming non-

We thank Maryellen MacDonald and Maggie Bruck for comments.

Correspondence concerning this article should be addressed to Mark S. Seidenberg, who is now at the Neuroscience Program, HNB 18B, University of Southern California, Los Angeles, California 90089-2520.

[1] Besner, Twilley, McCann, and Seergobin (1990) reported higher error rates in some of their simulations than we obtain using the same weights. The discrepancies probably relate to ambiguity concerning the correct pronunciations of nonwords such as PLOOD (Glushko, 1979), THA (Campbell & Besner, 1981), and GOLPH (McCann & Besner, 1987). First, it is not clear what the "regular" pronunciation of a nonword such as PLOOD is (like "good" or "food"?); second, there are differences in accent (the model was trained in Seidenberg's accent, which appears to differ from the one Besner et al. used in scoring the data). It is also questionable whether the same criteria were used in scoring the subject and model data. McCann and Besner scored as correct any pronunciation of a nonword that was consistent with the grapheme–phoneme correspondences given in the Hanna, Hanna, Hodges, and Rudorf (1966) list. The model's performance was assessed differently. Each nonword was assigned a single correct pronunciation, and Besner et al. determined how often this pronunciation provided the best fit to the computed output. Thus, the criteria used in scoring the model's performance apparently were more stringent.

Table 1
Errors on the Glushko (1979) Stimuli

Item	Target	Model's response	Comment
kead	kEd	ked	Regular pronunciation
plood	plUd	plud	Regular pronunciation
tost	t*st	tOst	Regular pronunciation
gomb	gom	gOm	Like *comb*
beash	bES	bAS	1 feature
brean	brEn	brAn	1 feature
breat	brEt	bret	1 feature
broff	br*f	bruf	1 feature
gomp	gomp	gump	1 feature
pote	pOt	pIt	1 feature
tife	tIf	tIv	1 feature
troad	trOd	trId	1 feature
kede	kEd	yEd	2 features
kull	k^l	y^l	2 features
kere	kEr	ZEr	2 features
kulp	k^lp	N^lp	2 features
pild	pild	pAld	2 features
[35 others correct]			

Note. The code used for representing phonemes is described in Seidenberg and McClelland (1989, p. 528).

words. Nonword performance therefore depends on vocabulary size. The model's performance reflects which words did or did not happen to be included in the training corpus. Given this theory of nonword naming, it would be anomalous if the model's performance did *not* differ from people's in the observed way.

The remaining question is whether a sufficiently large training corpus would allow the model to produce correct pronunciations for *all* nonwords, including ones such as PLAIE, DOWT, and TRUFE, which have no word-body neighbors. We cannot answer this question definitively without running a simulation, but the following points should be noted. For words, performance is primarily determined by frequency of exposure to a word itself and secondarily determined by the word-body neighbors. For nonwords, the primary impact comes from the word-body neighbors; there are secondary effects due to more remote neighbors. The nonword SOAT, for example, is affected by the SOA- words as well as the -OAT words. The same is true with human subjects; whereas only the word-body neighbors have discernible effects on the pronunciation of words, a much larger pool of neighbors affects nonword pronunciation (e.g., Taraban & McClelland, 1987). Two implications follow. First, the model should be able to piece together the pronunciations of nonwords such as TRUFE from exposure to the more remote neighbors. We already know this is true in some cases; the original simulation actually produced correct output for PLAIE, DOWT, and TRUFE and many other such nonwords. Second, limits on the size of the training corpus penalize the model even more than the Figure 1 data suggest. Many of the remote neighbors that will be relevant to stimuli such as McCann and Besner's (1987) are not in the corpus, either.

In summary, people are able to pronounce nonwords like DOWT on the basis of their knowledge of words; the model performs similarly, within the restrictions of the training corpus. It is important to understand the limits of the current simulation; it is also important to ask how well a person would pronounce nonwords if the person's vocabulary were limited to 2,900 words. According to our model, the answer is, probably pretty well for nonwords like MAVE and less well for items like FAIJE.

Wickelflaws

One other factor limits the model's performance on nonwords, the notorious Wickelfeatures, which we borrowed from Rumelhart and McClelland (1986). This representational scheme is an example of coarse, conjunctive coding (Hinton, McClelland, & Rumelhart, 1986). Some of the advantages of this type of coding were not fully captured by the Wickelfeature instantiation of the idea. This representation conjoins features of a phoneme with features of neighboring phonemes, but not with other features of the same phoneme. The net result is that the representation for one Wickelphone (e.g., /mAk/ as in *make*) is too similar to the representation of other, similar Wickelphones, such as /nAk/ in the nonword *nake*. The model then tends to produce output that differs from the correct response by just one feature. Besner et al. (1990) noted that the model makes a large number of errors that are one phoneme away from the correct answers; in fact, most errors are a single feature away (e.g., see Table 1). Approximately two thirds of the errors on the McCann and Besner (1987) stimuli are also one feature off.

We noted the inadequacy of the Wickelfeature scheme in our article. However, we were wrong to suggest that the use of this scheme did not contribute in any important way to the results. It did in fact have an effect: It artifactually increased the likelihood of single-feature errors. This limitation of the model will be addressed in the next generation of research, by developing an encoding scheme that increases the differentiation of similar Wickelphones (such as /mAk/ and /nAk/), which should reduce the number of single-feature errors. With a better phonological representation, only a modest increase in the size of the training corpus may result in significantly better nonword performance.

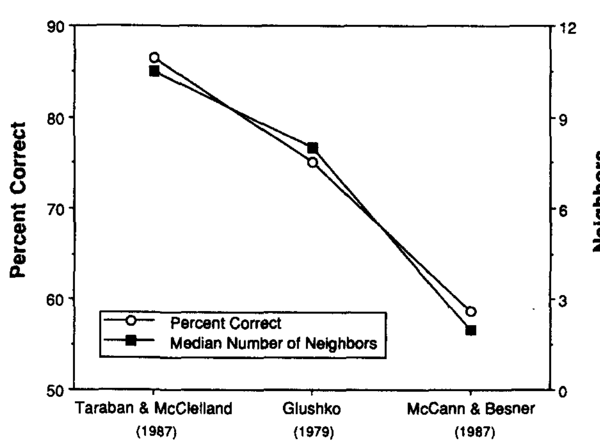

Figure 1. Relationship between neighborhood size and nonword naming accuracy.

In summary, given the dual limitations of the simulation—the size of the training corpus and the Wickelphonology—we think the model does surprisingly well on nonwords. The power of the learning rule is such that the model is able to pick up generalizations that support the correct pronunciation of many nonwords *despite* these limitations.

Other Naming Strategies

Even if the model is trained on a 30,000-word corpus with a perfect representation of phonology, it still will not simulate all aspects of human naming performance. People can and will use strategies that involve capacities that, although compatible with the model, are not within its immediate scope. This issue is illustrated by the Campbell and Besner (1981) study, in which subjects named nonwords such as THUF, THEIL, and THOVE. According to Besner et al. (1990), the model mispronounced more than 75% of these items. Again, our tally differs for reasons mentioned earlier, indicating that the model produced correct pronunciations for more than half the items. In any case, the model produces plausible output for most items, and where it makes clear errors, they are again one or two features from correct. Besner et al.'s main point is that whereas the subjects in the experiment pronounced the *th* as /T/ more than 80% of the time, the model produces the /T/ (as in *thin*) and /D/ (as in *thine*) pronunciations about equally often. However, the experiment involved repeated presentations of stimuli beginning with *th*. Repeating spelling patterns in words with differing pronunciations (e.g., HAVE–GAVE) produces interference in naming studies (Seidenberg, Waters, Barnes, & Tanenhaus, 1984). The question, then, is how subjects respond to a list of stimuli including items such as THAD, THA, THOVE, THAZ, and THEIL. One way to avoid the interference would be to always assign a single pronunciation to the digraph TH. One way to accomplish this would be to ignore the initial digraph, pronounce the word body, and add /T/, the higher frequency pronunciation of *th*-. This strategy would obviously result in an overwhelming proportion of /T/ responses. Our model does not attempt to simulate such effects, although modifying it to do so would be simple.

Monsell, Patterson, Tallon, and Hill (1989) reported an experiment in which they explicitly instructed subjects to use a similar strategy. Subjects were asked to "regularize" exception words such as PINT (i.e., to say /pint/). Again, the task could be performed by stripping off the initial consonant or consonant cluster, pronouncing the word body, and attaching the initial phoneme or phonemes. Again, although the current model does not simulate this strategy, modifying it to do so would be simple; if the model is given word bodies such as -INT in isolation, it reliably produces the regular pronunciations.

The Monsell et al. (1989) study shows that subjects can use a parsing strategy in naming familiar words. The extent to which this kind of strategy is used in naming nonwords is simply not known but will be important to investigate in the future. Confronted with a stimulus such as JINJE, subjects may find it more efficient to parse the stimulus into subcomponents that are easier to pronounce than the nonword as a whole. Such parsing strategies are clearly used by young readers in sounding words out, and the pathological condition termed *letter-by-letter reading* can be seen as an extreme application of the approach. There may be other strategies as well, such as drawing explicit analogies to particular lexical items.

In summary, it is inappropriate to assume that all aspects of nonword naming should be explained by a model that does not incorporate various other strategies. It is valid to ask whether the model can be extended in simple ways to deal with such phenomena; however, data such as Campbell and Besner's (1981) do not present a serious challenge. The model doesn't do pig Latin, either, but the relevant task-specific processes could certainly be added to what is already there.

Pseudohomophone Effects

Besner et al. (1990) asserted that the model cannot explain the pseudohomophone effects in the McCann and Besner (1987) and McCann, Besner, and Davelaar (1988) studies. This provides one of the main bases for their claim that the model needs lexical nodes. We have two reactions to arguments based on pseudohomophone effects. First, we do not think that these effects are inconsistent with the larger model, of which the implemented model is only a part. In the larger model, there are feedback connections from the phonological level to other parts of the system. These connections will tend to allow phonological representations to influence processing in other parts of the system. Because they have phonological representations that are the same as words, the processing of pseudohomophones should produce wordlike effects via these feedback connections. A second point, however, is that it remains uncertain whether genuine pseudohomophone effects exist and if so, under exactly which conditions they are obtained. The logic of the studies demands that pseudohomophones and nonpseudohomophones be equated in terms of other factors relevant to processing. Besner et al. described these stimuli as "tightly matched" (p. 434). They were equated in terms of word bodies and bigram frequencies. In Seidenberg and Waters's (1989) corpus of naming latencies for 3,000 words derived from 30 subjects, the correlation between mean bigram frequency and latency is −.07. McCann and Besner did not equate the stimuli in terms of initial phoneme. This factor accounts for approximately 10% of the variance in the Seidenberg and Waters corpus. Initial phonemes differ in frequency, ease of pronunciation, and acoustic properties that affect when responses are detected by the voice key. To give some hint of the problem, there are five words beginning with /v/ in the McCann and Besner stimuli; they yielded a mean naming latency of 750 ms, 145 ms longer than the remaining stimuli; all five are in the nonpseudohomophone controls.

McCann and Besner (1987) were aware of this problem and addressed it in a second experiment that involved two new sets of stimuli. These stimuli retained the initial phonemes from the pairs in the first experiment (e.g., BRANE–FRANE) but changed an internal vowel so that none of the items were pseudohomophones (e.g., BRONE–FRONE). Naming latencies for these new stimuli did not differ. However, a more direct comparison can be made with McCann and Besner's pseudohomophone and nonpseudohomophone stimuli. It is possible to select a subset of 54 pairs that are equated in terms of initial phoneme and length. This list includes all pairs that could be included while meeting these constraints. The net pseudohomophone effect in these stimuli is 10 ms, which does not approach significance, $t(1, 53) = -0.92$, $p > .35$. Thus, we question whether there is a pseudohomophone effect at all. Note that this account also explains why the effect for the original 80 pairs of stimuli remains even after the error scores from the model are partialed out. The error scores do not measure variability associated with the acoustic–phonetic properties of the initial phoneme. Hence, partialing them out does not remove the effect.

In summary, pseudohomophone effects have been problematic ever since Rubenstein, Lewis, and Rubenstein's (1971) original study and Clark's (1973) famous reanalysis of it. These effects are not inconsistent with the larger system of which our implemented model is but a fragment; however, the conditions under which such effects occur remain murky at best.

Lexical Decision

Besner et al. (1990) also questioned our account of the lexical decision task. They plotted the phonological error scores from the Waters and Seidenberg (1985) experiment and argued that words and nonwords are not discriminable on a phonological basis. They then asked why subjects in the experiment phonologically recoded. We explicitly cautioned against this use of the phonological error scores in our article (Seidenberg & McClelland, 1989, p. 529). The error scores are ones that we, the modelers, derive by comparing the computed output to the correct answer. The scores are then used to predict naming latencies. These error scores could not provide a basis for making lexical decisions, however, because there is no way for subjects to compute them without an external specification of the correct answers. Hence, the distributions of error scores in Besner et al.'s Figures 1 and 2 are relevant to their account of lexical decision, but not to ours.

They also reported the results of two experiments (Besner et al., 1990, Table 3). One was based on the observation that several of the strange words in the Waters and Seidenberg (1985) stimuli are homophones (e.g., AISLE–ISLE). Besner et al. (1990) thought this was important and repeated the study, replacing the problem words with nonhomophones. As in the original study, there was a regularity effect for low-frequency words. Hence, the study disconfirmed their hypothesis that the presence of homophones in the original stimuli had an impact on the results. In their second study, they replaced all of the strange words with homophones. The idea here was that the regularity effect might not require the presence of strange words. Besner et al. reported obtaining the effect. This experiment actually provides a very simple illustration of our lexical decision theory. Our theory says that phonological effects in lexical decision depend on how discriminable the words and nonwords are in terms of orthography. The more the distributions of orthographic error scores overlap, the bigger the phonological effect. For Besner et al.'s stimuli, the overlap is less than in Waters and Seidenberg's condition containing strange words but greater than in the condition in which the strange words were deleted. Most of the error scores for words and nonwords do not overlap; some do. Hence, the model predicts that phonology would only be consulted on a small proportion of trials. This might well produce an overall effect of regularity, but it would not be expected to generalize over items. That is exactly what Besner et al. found. The regularity effect was significant by subjects and not by items.

Besner et al. (1990) also questioned our account of the Waters and Seidenberg (1985) data (Seidenberg & McClelland, 1989, Figure 24). Our claim was that when the word stimuli consist of regular–exception words, the word–nonword decision can be based on orthographic information. Besner et al. noted that if the decision criterion for the data in Figure 24 is set to yield 5.2% errors on words, as in the experiment, the error rate for nonwords is 28%, which is too high. Setting the decision criterion in this way is overly restrictive: The error rate in the experiment is an average based on 28 subjects, whereas the simulation data represent 1 subject. Each run of the simulation produces different error scores because words are sampled randomly during the training phase; this effect is compatible with the small individual differences between subjects that are observed. Hence, the decision criterion for the Figure 24 data does not have to yield 5.2% errors. For these data, we predict that decisions would primarily be based on orthographic information, given the small amount of overlap in the regular–exception word and nonword distributions. Phonology would be used only where the words and nonwords overlap. Given the small number of items in question, no overall effect of phonological regularity is predicted to obtain, as in the experiment.

The same account of lexical decision performance applies to a third experiment reported by Besner et al. (1990), which concerned pseudohomophones. McCann et al. (1988) reported longer lexical decision latencies for pseudohomophones compared with nonpseudohomophones. The stimuli in this experiment are not equally wordlike, as indicated by the fact that the pseudohomophones produce smaller orthographic and phonological error scores. This is another consequence of failing to equate the stimuli in terms of initial phoneme. In our 1989 article, we described an experiment with two new sets of nonwords, equated in terms of error scores, in which pseudohomophones did not yield longer latencies than nonpseudohomophones. Besner et al. raised various methodological objections to this study; however, when they repeated the study with minor modifications of the stimuli, they replicated our result: no pseudohomophone effect on lexical decision latencies. As in our study, pseudohomophones yielded more errors. These errors are wholly consistent with our account of lexical decision. Most of the word and nonword stimuli in these experiments can be differentiated on the basis of orthography. If the pseudohomophones and nonpseudohomophones are equated in terms of orthographic properties, they will be equally difficult to discriminate from words, predicting no overall latency difference between the two types of nonwords. A few of the pseudohomophones (e.g., GANE and FEAL) are difficult to discriminate from words on an orthographic basis. Our theory suggests that subjects will have to consult phonological information on these trials. Phonologically recoding a pseudohomophone such as GANE may result in activation of the meaning associated with the homophonous word GAIN. On a small proportion of trials, this activated semantic information is sufficient to cause a false-positive response (see Van Orden, 1987, for similar results). Thus, the model correctly predicts no overall pseudohomophone effect when the two types of nonwords are equally wordlike; it also makes the more subtle prediction that false-positive responses should be more likely to occur when pseudohomophones are difficult to discriminate from words on an orthographic basis. These results are consistent with our view that decision criteria vary in response to properties of the stimuli; they provide little support for Besner et al.'s broader claim that pseudohomophones are always processed by accessing the base words from which they are derived.

In summary, there is nothing about Besner et al.'s (1990) experiments that is inconsistent with our account of lexical decision. In questioning why subjects phonologically recode in lexical decision, we think that Besner et al. overlooked an obvious possibility, that subjects may do so because they are using phonology to access meaning, which is useful in discriminating words from nonwords. It will take further research to establish whether it is phonology alone, or semantics alone, or both that contribute to the decision process under particular stimulus conditions.

Other Concerns

Two of Besner et al.'s (1990) other concerns should be addressed. First, they remarked several times that the error scores fail to account for much of the variance in response latencies. The Seidenberg and Waters (1989) corpus of naming latencies for 3,000 words provides a rigorous basis for assessing this claim. The corpus contains statistics concerning a broad range of measures thought to influence lexical processing (e.g., frequency, Coltheart's N, and bigram frequency), as well as error scores from the model. Entering phonological error score first in a stepwise regression yields an R of .29. Adding the factor length in letters increases the multiple R to .41. Error scores do not reflect length because of the way the model is set up; each trial begins with a word being encoded as a pattern of activation over the orthographic units. Thus, the model does not simulate letter-recognition processes that depend on length. Once length and error score are entered in the regression, the effects of other factors such as Coltheart's N, bigram frequency, and Kucera-Francis frequency are almost entirely eliminated. The only other factor that matters is initial phoneme, which brings the multiple R up to approximately .7. In summary, although these data are still being analyzed, it is clear that the error scores account for significant amounts of variance, certainly more than some other factors that are standardly manipulated in experiments. Clearly, part of the unexplained variance may be due to the fact that error scores do not reflect the words that were not in the training corpus.

Finally, Besner et al. (1990) criticized our claims about a single route for naming. We actually discussed two routes: the one we implemented, and a second route through meaning. The single-route idea was that the process we implemented could generate correct output for regular and irregular words and nonwords. This contrasted with the nearly universal prior intuition that at least two processes would be necessary to accommodate all these cases. Once it is acknowledged that a single process can generate correct output in all these cases, we think it is important to consider the division of labor between the routes. We agree with the insight of dual-route modelers that these processes jointly support performance; we merely disagree on what kinds of knowledge they involve and how they work.

Conclusions

As we stated in the 1989 article, our model is limited, and there is plenty of room for further development. Besner et al.'s (1990) idea that there are entries for individual lexical items is an interesting one that might be developed as part of an explicit alternative model. It would certainly be impressive if such a model could show how knowledge of spelling–sound correspondences is acquired and represented in memory, simulate in quantitative detail the results of 20 or 25 experiments, generate new predictions that are confirmed in subsequent studies, provide an account of individual differences in reading skill, capture basic facts about the acquisition of this skill, show how differences among orthographies influence processing, generate several novel hypotheses about the bases of developmental dyslexia, perform like several acquired dyslexic patients in the literature when damaged, and provide a theory of how the lexical decision task is performed. It would also be impressive if the model could generate the correct pronunciations of nonwords such as JINJE on the basis of exposure to 2,900 words. A model of this sort would be a worthy candidate for consideration as an alternative to ours. If that model could also produce a pseudohomophone effect—and there were a real pseudohomophone effect in the behavioral data—we would have to conclude that it is indeed the better model.

References

Besner, D., Twilley, L., McCann, R. S., & Seergobin, K. (1990). On the association between connectionism and data: Are a few words necessary? *Psychological Review, 97,* 432–446.

Campbell, R., & Besner, D. (1981). This and thap: Constraints on the pronunciation of new, written words. *Quarterly Journal of Experimental Psychology, 33A,* 375–396.

Glushko, R. (1979). The organization and activation of orthographic knowledge in reading aloud. *Journal of Experimental Psychology: Human Perception and Performance, 5,* 674–691.

Hanna, P. R., Hanna, J. S., Hodges, R. E., & Rudorf, E. H. (1966). *Phoneme-grapheme correspondences as cues to spelling improvement.* Washington, DC: U.S. Department of Health, Education, and Welfare.

Hinton, G. E., McClelland, J. L., & Rumelhart, D. E. (1986). Distributed representations. In D. E. Rumelhart & J. L. McClelland (Eds.), *Parallel distributed processing* (Vol. 1, pp. 77–109). Cambridge, MA: MIT Press.

McCann, R., & Besner, D. (1987). Reading pseudohomophones: Implications for models of pronunciation assembly and the locus of word-frequency effects in naming. *Journal of Experimental Psychology: Human Perception and Performance, 13,* 14–24.

McCann, R., Besner, D., & Davelaar, E. (1988). Word recognition and identification: Do word-frequency effects reflect lexical access? *Journal of Experimental Psychology: Human Perception and Performance, 13,* 693–706.

Monsell, S., Patterson, K. E., Tallon, J., & Hill, J. (1989, January). *Voluntary surface dyslexia: A new argument for two processes in reading?*

Paper presented at the meeting of the Experimental Psychology Society, London.

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

Rumelhart, D. E., & McClelland, J. L. (1986). On learning the past tenses of English verbs. In J. L. McClelland & D. E. Rumelhart (Eds.), *Parallel distributed processing* (Vol. 2, pp. 216–271). Cambridge, MA: MIT Press.

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

Seidenberg, M. S., & Waters, G. S. (1989, November). *Naming words aloud: A mega-study.* Paper presented at the meeting of The Psychonomic Society, Atlanta, GA.

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

Taraban, R., & McClelland, J. L. (1987). Conspiracy effects in word recognition. *Journal of Memory and Language, 26,* 608–631.

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

Waters, G. S., & Seidenberg, M. S. (1985). Spelling-sound effects in reading: Time course and decision criteria. *Memory & Cognition, 13,* 557–572.

Received March 6, 1990
Accepted March 6, 1990 ■