Journal of Experimental Psychology:
Human Perception and Performance
1994, Vol. 20, No. 6, 1177–1196

Copyright 1994 by the American Psychological Association, Inc.
0096-1523/94/$3.00

Nonword Pronunciation and Models of Word Recognition

Mark S. Seidenberg, David C. Plaut, Alan S. Petersen, James L. McClelland, and Ken McRae

Nonword pronunciation is a form of generalization behavior that has been at the center of debates about models of word recognition, the role of rules in explaining behavior, and the adequacy of the parallel distributed processing approach. An experiment yielded data concerning the pronunciation of a large corpus of nonwords. The data were then used to assess 2 models of naming: a model developed by D. C. Plaut and J. L. McClelland (1993), which is similar to the one described by M. S. Seidenberg and J. L. McClelland (1989) but uses improved orthographic and phonological representations, and the grapheme–phoneme correspondence rules of M. Coltheart, B. Curtis, P. Atkins, and M. Haller's (1993) dual-route model. Both models generate plausible nonword pronunciations and match subjects' responses accurately. The dual-route model does so by using rules that generate correct output for most words but mispronounce a significant number of exceptions. The parallel distributed processing model does so by finding a set of weights that allow it to generate correct output for both "rule-governed" items and exceptions. Some ways in which the two approaches differ and other issues facing them are also discussed.

The task of reading words and nonwords aloud has played a central role in the development of models of word recognition. Reading makes use of knowledge concerning the correspondences between the orthographic and phonological forms of words. This information is used in recognizing words and pronouncing them aloud (see Seidenberg, in press, for a review). Current models differ in their assumptions about how this knowledge is acquired, represented, and used. Dual-route models assume that there are separate lexical and sublexical procedures for generating pronunciations (for an overview, see Patterson & Coltheart, 1987; for critiques, see Humphreys & Evett, 1985, and Van Orden, Pennington, & Stone, 1990). The specific version of the dual-route model developed by Coltheart and his colleagues (Coltheart, 1978, 1987; Coltheart, Curtis, Atkins, & Haller, 1993) assumes that in alphabetic writing systems, knowledge of the correspondences between orthography and phonology is represented in terms of rules translating graphemes into phonemes. The rules are used in naming words whose pronunciations they correctly specify (sometimes called "regular" words). Words whose pronunciations violate the rules ("exceptions" such as HAVE, PINT, and GONE) must be pronounced by means of a separate lexical (or word-specific) pronunciation mechanism.

The fact that people are able to pronounce novel, nonword letter strings such as NUST and MAVE has been taken as further evidence for the hypothetical grapheme–phoneme correspondence (GPC) rules. The ability to generalize has provided the classic source of evidence for mental rules. For example, the fact that the children in Berko's (1958) famous experiment correctly used novel forms such as WUGS ("this is a WUG, here are two __.") was taken as evidence that they had learned a rule of plural formation. By the same reasoning, people's ability to pronounce NUST as a rhyme of MUST and DUST has been taken as indicating that they have acquired GPC rules. These rules play other roles in the dual-route model as well. Acquisition of the rules is thought to be an early step in learning to read; poor knowledge of the rules is associated with failures to acquire age-appropriate reading skills, and the acquired reading disorder phonological dyslexia is thought to reflect the loss of these rules due to brain injury (Castles & Coltheart, 1993; Coltheart, 1987). Coltheart et al. (1993) recently described an algorithm for inducing a set of grapheme–phoneme rules and using them to pronounce words and nonwords. The rules generate correct output for about 78% of the 2,897 monosyllabic words in a corpus developed by Seidenberg and McClelland (1989); they also generate plausible output for nonwords.

Connectionist (or "parallel distributed processing") models have challenged this view by providing an alternative to the assumption that regularities can only be represented in terms of rules. Networks using distributed representations, weighted connections between units, and error-minimization learning algorithms can encode both "rule-governed" cases and "exceptions." For example, simple feedforward networks such as the ones described by Sejnowski and Rosenberg (1987) and Seidenberg and McClelland (1989)

Mark S. Seidenberg and Alan S. Petersen, Neuroscience Program, University of Southern California; David C. Plaut and James L. McClelland, Department of Psychology, Carnegie Mellon University; Ken McRae, Department of Psychology, University of Western Ontario, London, Ontario, Canada.

This research was supported by National Institute of Mental Health (NIMH) Grant MH47566, National Science Foundation Grant BNS 88-12048, and NIMH Career Development Award MH00385. We thank Max Coltheart for providing his rule algorithm and for help in using it, Marni Harrington for assistance with the experiment, and Maryellen MacDonald for comments on a draft of this article.

Correspondence concerning this article should be addressed to Mark S. Seidenberg, Neuroscience Program, University of Southern California, Los Angeles, California 90089-2520. Electronic mail and requests for the data for the 590 nonword corpus described in this article may be sent to marks@neuro.usc.edu.

took orthographic patterns as input and produced phonological codes as output. The models learned to produce correct output for large vocabularies of words, including regular–irregular pairs such as GAVE–HAVE and BONE–GONE.

These models also provide an alternative view of the bases of generalization: The information concerning spelling–sound correspondences derived from exposure to actual words and encoded by the weights in such networks is also used in generating pronunciations for unfamiliar stimuli. For example, given simple nonwords on which it had not been trained such as NUST or FIKE as input, Seidenberg and McClelland's model (1989) produced the pronunciations that people generate as output. The model thereby challenged what Seidenberg (1989) called the "central dogma" of dual-route models: that separate mechanisms (rules and lexical lookup) are needed for pronouncing nonwords and exceptions. Seidenberg and McClelland's model replaced the two naming mechanisms that were integral to the dual-route model with a single mechanism using weights on connections between units.

As Besner, Twilley, McCann, and Seergobin (1990) noted, the model did not perform as well as people on nonwords, especially ones such as JINJE, FAIJE, and TUNCE that contain unusual spelling patterns. In light of the fact that nonword pronunciation has been taken as evidence for pronunciation rules, this defect is potentially important. Deviations in the model's nonword performance could be taken as evidence that rules are needed in order to achieve humanlike performance, as Besner et al. (1990) and Coltheart et al. (1993) concluded. They have further claimed that basic limitations on the capacities of neural networks preclude their being able to generate correct output for both regular and irregular words while maintaining good generalization. Pinker and his colleagues (Pinker, 1991; Prasada & Pinker, 1993) have drawn the same conclusion from connectionist models of the past tense.

The limitations of simulation models need to be evaluated carefully, however. All such models are restricted in scope, ensuring that their behavior will deviate from that of people at some level of precision. The fact that a model's performance differs from people's does not itself reveal whether the limitations derive from defects in the theory on which the model is based or from details of the implementation that are not theory relevant. The behavior of the model—how it both corresponds to and deviates from human behavior—needs to be interpreted in light of the principles that govern its performance. This can be achieved by drawing on foundational research on the properties of such networks, by performing careful analyses of the network, and by experimenting with other networks designed to solve similar kinds of problems (see Plaut & Shallice, 1993, and Seidenberg, 1989, 1993, for discussions).

Drawing on the first two of these sources of information, Seidenberg and McClelland (1990) noted that the performance of their model was limited by at least two major factors: the size of the training corpus and the properties of the phonological representation that were used. The training corpus was 2,897 monosyllabic words, which is an order of magnitude smaller than a skilled reader's vocabulary. The corpus represented a sample out of the space of orthographic–phonological correspondences in English. The model performed well on nonwords that included these correspondences. Increasing the size of the sample would result in better coverage of this space, improving the model's performance on generalization trials. These observations suggest that good performance on simple nonwords such as FIKE and poorer performance on difficult nonwords such as FAIJE is what might be expected of a person who had acquired only a relatively small vocabulary. This could then be taken as a good example of a practical limitation on an implementation whose theoretical implications are minimal.

Coltheart et al. (1993) have challenged these observations about the effects of corpus size. Their algorithm induced a set of pronunciation rules on the basis of exposure to the same 2,897-word corpus that Seidenberg and McClelland (1989) used. The rules yielded significantly better performance on difficult nonwords such as FAIJE. Hence, Coltheart et al. concluded that the flaws in the Seidenberg and McClelland model could not be due to simply the size of the training corpus. This argument is not valid, however. The fact that rules sufficient to support the pronunciation of difficult nonwords can be induced from the 2,897-word corpus is orthogonal to the effects of corpus size on the network. The two models are being asked to solve very different problems and are affected by different factors. The Coltheart et al. algorithm has to induce rules that generate accurate output for nonwords but are allowed to err on many words because these items can be treated as exceptions and pronounced by a separate lexical processing mechanism. The network, by contrast, must perform the more difficult task of generating correct output for both regular and irregular words as well as nonwords. A 2,897-word vocabulary might be sufficient for the first task but not the second.[1]

The second limitation noted by Seidenberg and McClelland (1990) concerned the phonological representation that was used. This representation was constructed according to basic principles concerning distributed representations that were relevant to the theoretical claims being offered. The main principle was that words with similar phonological codes should activate similar patterns over the phonological units, and analogously for the orthographic units. Constructed in this way, the representation was sufficient to allow exploration of many issues concerning the mapping between orthography and phonology, but it was not a complete phonological system. Its defects became apparent at the limits of the model's performance: pronouncing nonwords such as JINJE and FAIJE. In such cases the model

[1] Note that the claim is not that performance would improve because all of the correspondences contained in nonwords such as FAIJE or TUNCE would necessarily be found in the larger corpus. Looking at the Kuçera and Francis (1967) corpus, for example, there are no -AIJE words at all. Some additional low-frequency correspondences will be found in a larger corpus, however (e.g., it might include DUNCE, which was not in Seidenberg and McClelland's, 1989, list), as would other correspondences that provide information relevant to patterns such as -AIJE by containing parts of them.

sometimes produced small deviations from correct targets. Plaut, McClelland, Seidenberg, and Patterson (1994) pointed out that the two issues discussed by Seidenberg and McClelland (1990)—the size of the corpus and the adequacy of the phonological representation—are not independent. Achieving an adequate level of performance while using a relatively coarse phonological representation may require exposure to a broader range of exemplars. Conversely, using a phonological representation that captures more of the relevant distinctions may afford this level of performance with a smaller corpus.

In summary, the issues that have arisen in connection with the task of naming nonwords aloud carry broader implications concerning the adequacy of connectionist models, the role of rules in human behavior, and the bases of the capacity to generalize. In this article we provide new information bearing on these issues. We first describe the results of a large-scale behavioral experiment on nonword naming. This experiment provides a rich set of data that is then used to assess two models: the parallel distributed processing (PDP) model described by Plaut and McClelland (1993), which used improved orthographic and phonological representations, and the Coltheart et al. (1993) GPC rules. The principal goal of these analyses was to assess the validity of the claim that connectionist models cannot generate correct output for both nonwords and exceptions at a sufficiently high level of accuracy and the corollary that two mechanisms, one of which uses pronunciation rules, are necessary.

The Experiment

Method

Subjects. The subjects were 24 McGill University undergraduates, native speakers of Canadian English, who were paid for participation.

Stimuli. The stimuli were monosyllabic nonwords created from 590 different word bodies (rimes) found in the 2,897 word corpus used by Seidenberg and McClelland (1989). The word bodies (e.g., -ANT, -OWN, -ANCE) were paired with onsets (single consonants or consonant clusters) to form nonwords. The entire set of stimuli is listed in the Appendix. The data for 10 additional items were deleted because of experimenter errors that resulted in missing scores. The items were divided into three randomized lists. Each subject was presented with all lists, with order of lists counterbalanced across subjects. There was also a list of 12 practice items, using word bodies that did not occur in the test stimuli.

Procedure. Stimuli were presented one at a time with a 2-s intertrial interval on an IBM PS2 Model 80 PC in a dimly lit room. Subjects were informed that the stimuli were nonwords and told to pronounce them as if they were words. They were given the practice list with feedback about their performance and then the three lists of test stimuli without any feedback. Subjects sat at a comfortable distance from the computer and spoke their responses into a microphone connected to a voice key interfaced to the computer. The experimenter, a speaker of Canadian English, recorded subjects' pronunciations by hand using the phonetic transcription given in the Appendix. With short breaks between blocks, the experiment took about 1 hr to run.

Results

Fewer than 1% of the trials were lost due to equipment malfunctions. Naming latencies more than 2 *SD*s above a subject's mean (1.3%) were replaced with the 2-*SD* value. Data were analyzed in terms of the number of pronunciations per nonword and the latencies associated with different pronunciations.

The nonwords varied in terms of the number of pronunciations they elicited across subjects. As Figure 1 indicates, subjects produced a single pronunciation for 34.7% of the items; another 45.9% elicited two pronunciations, 16.9% three pronunciations, and 2.5% four or more pronunciations. The last group included many pronunciations that were produced by only 1 or 2 subjects. These low-frequency responses consisted of both uncommon but possibly intended pronunciations and true errors. Because the line between these two types of responses is unclear (e.g., is /brat/ a mispronunciation of BREAT or a pronunciation by analogy to YEAH?), we categorized them together as "other" responses.[2]

Data concerning the frequencies of the alternative pronunciations indicate that although many items yielded multiple pronunciations, subjects nonetheless showed considerable agreement. The most common pronunciation for each nonword accounted for 83.7% of all responses. The second most common pronunciation accounted for an additional 9%. Thus, the two most common pronunciations accounted for more than 90% of the responses.

Naming latencies varied as a function of two factors: (a) the number of pronunciations that the item generated across subjects and (b) the frequency with which the pronunciation was generated across subjects. Table 1 provides data indicating how the generation of the most common pronunciation was affected by the availability of alternative pronunciations. Two hundred six items yielded only a single pronunciation each, with a mean latency of 656 ms. Two hundred sixty-nine items yielded two pronunciations across subjects. The mean naming latency for the dominant pronunciation, generated by 83.4% of the subjects, was 692 ms. For the 100 items that generated three pronunciations, the dominant pronunciation accounted for 60.7% of the responses, with a mean latency of 744 ms. Thus, the latency to produce the dominant pronunciation increased as a function of the number of alternative pronunciations. According to the dual-route model, subjects generate nonwords by applying GPC rules. The fact that different pronunciations are generated across subjects can be explained by assuming that they have slightly different rule sets. However, the data indicate that generating the most common, "rule-governed" pronunciation of a nonword was affected by the existence of alternative pronunciations. This effect reflects the degree of consistency in the mapping between spelling and pronunciation. As Glushko (1979) originally noted, the pronunciation of a word or nonword is affected by the degree of consistency among the pronunciations of its neighbors.

[2] The key to the pronunciation symbols can be found at the end of the Appendix.

Another way to observe this consistency effect is to look at the latencies associated with the first, second, and third most common pronunciations (see Table 2). Consider first the most common pronunciation for each of the 590 items. The mean latency for these pronunciations was 689 ms. For 384 items, subjects produced two or more pronunciations. Looking at the second most common pronunciation for these items, the mean latency was 715 ms. Similarly, for the 115 items that yielded three or more pronunciations, the mean naming latency for the third most common pronunciation was 783 ms. These data indicate that subjects' pronunciations were influenced by their knowledge of alternative pronunciations. Subjects not only generated atypical pronunciations on some trials, but they took longer to do so, suggesting that pronunciations were slowed by competition from neighbors with other pronunciations (see also Taraban & McClelland, 1987).

In summary, the experiment yielded orderly nonword naming data that replicate and extend earlier findings. The data are being made available electronically and can be used by other researchers in testing additional hypotheses and models.

Assessment of the Two Models

We now consider how subjects' performance relates to that of the two models. We first provide brief descriptions of the models and how the data were scored.

The Plaut and McClelland (1993) Model

The Plaut and McClelland (1993) model is a network that like Seidenberg and McClelland's (1989) model, performed the task of generating phonological codes from orthographic input. The principal difference between the two models relates to the introduction of improved orthographic and phonological representations. Other modifications that took advantage of the availability of increased computational resources as well as progress since 1989 in neural network theory were also introduced.

Network architecture. The architecture of Plaut and McClelland's (1993) network is shown in Figure 2. The network has three layers of units: 108 grapheme units, 100 hidden units, and 57 phoneme units. The grapheme units are fully connected to the hidden units, and the hidden and phoneme units are fully inter- and intraconnected. Each connection has a positive or negative real-valued weight that changes over the course of learning. In addition, as is standard in connectionist modeling, each hidden and phoneme unit has a *bias* value that determines the unit's default tendency to be on or off in the absence of contributions from other units (see Bechtel & Abrahamsen, 1991, for a discussion). This bias can be implemented as the weight on an additional connection from an extra unit that is always active. In this way, bias values can be learned in exactly the same way as all other weights in the network. Including the bias connections, the network has a total of 23,203 connections.[3]

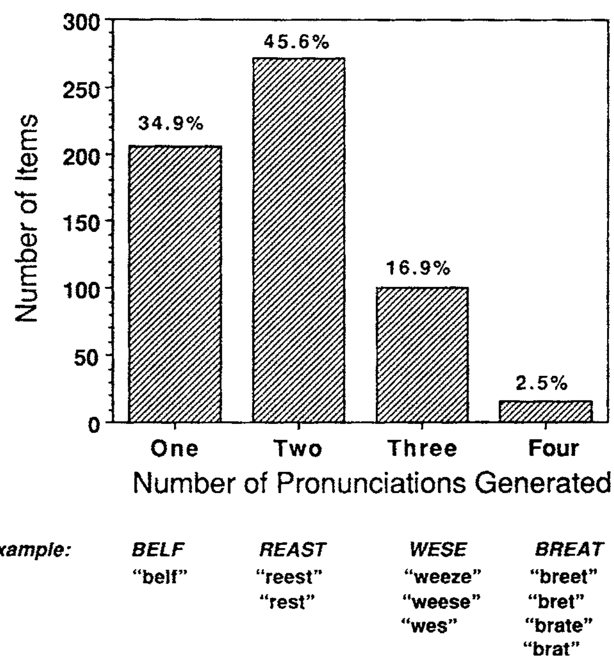

Figure 1. Number of pronunciations generated by each nonword.

Table 1
Naming Latency as a Function of Number of Alternative Pronunciations

No. given and variable	Pronunciations: First	Second	Third	Model
One				
RT	656 (3.6)			2.503 (.032)
%	97.1			
n	206			
% Errors	2.9			
Two				
RT	692 (4.1)	700 (8.6)		2.699 (.031)
%	83.4	12.1		
n	269			
% Errors	4.5			
Three				
RT	744 (8.1)	753 (10.7)	787 (18.7)	2.901 (.053)
%	60.7	23.4	10.4	
n	100			
% Errors	5.6			

Note. No. given refers to number of pronunciations given for a nonword across subjects. For example, 269 items yielded two pronunciations, with mean naming latencies for each pronunciation as indicated. Model column indicates mean settling time and standard error (in parentheses). RT = reaction time (naming latency in milliseconds) and standard error (in parentheses); % = percentage of correct responses; *n* = number of items.

[3] The grapheme units are not interconnected because, in the simulation, their states are completely determined by the input to the network. In a more realistic implementation of the visual processes involved in reading, grapheme units would become active gradually over time on the basis of more primitive visual information represented at even earlier levels of the system.

Each unit in the network has an activity level or state, ranging between 0.0 and 1.0, and each connection from one unit to another has a real-valued weight that can be positive or negative. In a standard connectionist network, the state s_j of each unit j is a smooth, nonlinear (logistic) function $\sigma(\cdot)$ of its summed input x_j from other units:

$$x_j = \sum_i s_i w_{ij} \tag{1}$$

$$s_j = \sigma(x_j) = \frac{1}{1 + \exp(-x_j)}, \tag{2}$$

where w_{ij} is the weight from unit i to unit j and $\exp(\cdot)$ is the exponential function. In the current network, the states of units change gradually over time. Specifically, the new state of unit j at time $t + \tau$, $s_j^{[t+\tau]}$, is a weighted average of its current state at time t and the state dictated by its summed input:

$$x_j^{[t+\tau]} = \sum_i s_i^{[t]} w_{ij} \tag{3}$$

$$s_j^{[t+\tau]} = \tau\,\sigma(x_j^{[t]}) + (1 - \tau)s_j^{[t]}, \tag{4}$$

where τ is the weighting proportion that determines how gradually the states of units change and $\sigma(\cdot)$ is the standard nonlinear unit function shown in Equation 2. For $\tau = 1$, the second term in Equation 4 is zero and the units function as in a standard network (cf. Equation 2). As τ approaches zero, the network can be viewed as an increasingly close discrete approximation to a system that is continuous in time.

Representations. Letter strings are represented by specific patterns of activity over the input graphemic units, and monosyllabic pronunciations are represented over the output phonemic units. The orthographic and phonological representations in the model were designed to address some of the limitations of earlier approaches. The representations used by Seidenberg and McClelland (1989) suffered what Plaut et al. (1994) termed a "dispersion" problem. The information that was relevant to a particular phoneme in a particular position was dispersed over different units (i.e., over the different Wickelphone units that the phoneme activated).[4] One type of representation that does not suffer from this problem involves using position-specific representations (McClelland & Rumelhart, 1981). However, this requires repeating phonemes at every position, which introduces other problems. For example, the representation of any given word becomes extremely sparse, and there is nothing relating occurrences of the same phoneme across different positions (thus, the *b*s in BAT and TAB would be treated as unrelated). The representation used here is a compromise that involves syllabic rather than phonemic positions: onset, nucleus, and coda. These syllabic units play a central role in phonological theories (e.g., Selkirk, 1982), speech errors (e.g., Dell, 1986), and the acquisition of reading skill (Treiman, 1992). Each unit represents a particular grapheme or phoneme within one of these positions (see Table 3). Graphemes can be either single letters or multiletter "relational units" (Venezky, 1970) that have a specific phonological correspondence (e.g., PH → /f/). As the parsing of letter strings into graphemes is ambiguous in the general case (e.g., TOPHAT VS. CELLOPHANE), all possible graphemes within a string are activated, including the components of multiletter graphemes (e.g., P, H, and PH).

The phonotactic constraints on which sequences of phonemes create well-formed monosyllabic pronunciations in English dictate that the identities of phonemes within each cluster are sufficient to determine their ordering (as reflected in Table 3). The only violations to this generalization involve /s/ and /p/, /t/, or /k/ in the coda (e.g., CLASP vs. LAPSE). Three additional units—/ps/, /ks/, and /ts/—are required to handle these cases. The treatment of these units is analogous to that of multiletter graphemes in that their presence is indicated by the simultaneous activation of the affricate unit and the units representing its component phonemes. Additional motivation for these units is provided by the observation that these combinations are sometimes treated as single phonemes, called affricates, and sometimes written with single letters (e.g., English X, Greek Ψ). Analogous orthotactic constraints ensure that the ordering of letters within a string is unambiguously represented by the identities of its graphemes within each orthographic consonant and vowel cluster.

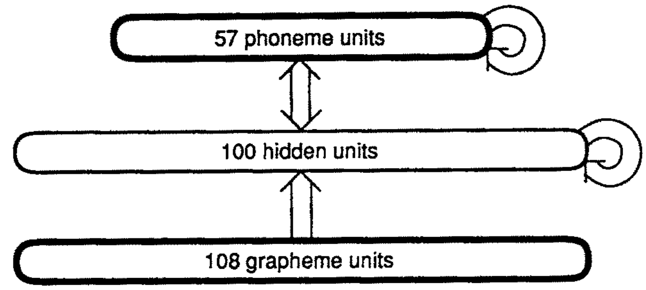

Figure 2. Architecture of the Plaut and McClelland (1993) model.

Table 2
Data as a Function of Pronunciation Frequency

Pron	*n*	Latency (ms)
1	590	689
2	384	715
3	115	783
4	15	752

Note. Pron refers to the frequency with which a pronunciation was given. Pron 1 is based on the most common pronunciation for every nonword. Pron 2 is based on the second most common pronunciation for the 384 items that yielded two or more pronunciations, and so on.

[4] Seidenberg and McClelland's (1989) phonological representation was the "Wickelphonology" developed by Rumelhart and McClelland (1986). The orthographic representation was developed specifically for their model.

Table 3
Orthographic and Phonological Representations

Syllabic position	Type of unit
Phonological[a]	
Onset	s b p d t g k f v z T D S Z l r w m n h y
Vowel	a @ e i o u A E I O U W Y ^
Coda	r l m n N b g d ps ks ts s f v p k t z S Z T D
Orthographical	
Onset	Y S P T K Q C B D G F V J Z L M N R W H U CH GH GN GU PH PS QU RH SH TH TS WH
Vowel	E I O U A Y AI AU AW AY EA EE EI EU EW EY IE OA OE OI OO OU OW OY UE UI UY
Coda	H R L M N B D G C X F V J S Z P T K BB CH CK DD DG FF GG GH GN GU KS LL NG NN PH PP PS QU RR SH SL SS TCH TH TS TT ZZ E ES ED

Note. Symbols are from "Generalization with componential attractors: Word and nonword reading in an attractor network" (pp. 824–829), by D. C. Plaut and J. L. McClelland, 1993, in *Proceedings of the Fifteenth Annual Conference of the Cognitive Science Society*, Hillsdale, NJ: Erlbaum. Copyright 1993 by Erlbaum. Adapted by permission.
[a] /a/ in POT, /@/ in CAT, /e/ in BED, /i/ in HIT, /o/ in DOG, /u/ in GOOD, /A/ in MAKE, /E/ in KEEP, /I/ in BIKE, /O/ in HOPE, /U/ in BOOT, /W/ in NOW, /Y/ in BOY, /^/ in CUP, /N/ in RING, /S/ in SHE, /Z/ in BEIGE, /T/ in THIN, /D/ in THIS.

Training procedure. The training corpus consisted of the 2,897 monosyllabic words in the Seidenberg and McClelland (1989) corpus, augmented with 101 words missing from that corpus but used as stimuli in various behavioral studies. The corpus also included patterns consisting of each grapheme in isolation and the corresponding phonemes. The reasoning here was that the model should be explicitly trained on these correspondences because children typically are taught them in the course of learning to read. As it turns out, omitting this training on isolated GPCs has little impact on network performance (see Plaut et al., 1994).[5]

An input string is presented to the network by clamping the states of the appropriate grapheme units. The network is given a fixed amount of time to process the input (t = 0.0–3.0), during which the units change their states gradually according to Equations 3 and 4 (with $\tau = 0.5$). The network's performance at each point in time was measured by the cross-entropy (Hinton, 1989; Kullback & Leibler, 1951) of the phoneme units' activities with their target activities for this input:

$$C^{[t]} = -\sum_j t_j^{[t]} \log_2(s_j^{[t]}) + (1 - t_j^{[t]}) \log_2(1 - s_j^{[t]}), \qquad (5)$$

where j indexes phoneme units and t_j is the target for each. Like the more standard total sum of squared error measure, cross-entropy is a measure of the difference between the phoneme states generated by the network and their correct (target) states for each word. Cross-entropy is a more appropriate error measure when the state of each output unit can be interpreted as the probability that a particular hypothesis is true (Rumelhart, Durbin, Golden, & Chauvin, in press). This applies to the current task, as each phoneme unit corresponds to the hypothesis that a particular phoneme is present in the network's response. More formally, the states of phoneme units, when interpreted as independent probabilities, define a probability distribution over all possible responses. The targets of phoneme units define another probability distribution (although, if all targets are either 0 or 1, the distribution simply assigns a probability of 1 to a particular response). Cross-entropy measures the information-theoretic distance between these two probability distributions (Kullback & Leibler, 1951).

In order to encourage the network to be correct as quickly as possible, the magnitude of the cross-entropy error was weighted to gradually increase over time. Specifically, the weighting was set to 0.0 until $t = 1$ and then linearly increased, reaching 1.0 at the end of settling ($t = 3$). Also, the network was halted and received no more error once it succeeded in activating all phonemes to within 0.2 of their correct values. Words satisfying this criterion were guaranteed to be pronounced correctly given the procedure for generating responses from phonological activity, described below.

The weights on all connections were initialized to small random values. As a result, at the beginning of training, the phoneme activations generated by the network for each word were much different from the correct activations for the word (i.e., the cross-entropy error was high). A version of backpropagation through time (Rumelhart, Hinton, & Williams, 1986; Williams & Peng, 1990) adapted for continuous units (Pearlmutter, 1989) was used to compute how to change each weight so as to reduce the error on each word (see Plaut et al., 1994, for details). The procedure is

[5] Strictly speaking, children cannot be taught the correspondences of nonfricative consonants in isolation because the pronunciation of such consonants must be followed at least by a neutral vowel (/^/ in our notation; see Table 3). However, because the variation across the pronunciations of such minimal syllables is attributable almost entirely to the consonant itself, the effects of training directly on such syllables would be equivalent to training on the isolated correspondences of the consonants themselves.

essentially the same as standard backpropagation except that rather than receiving error in a single backward pass through the network, units gradually accumulate error in the same way as they accumulate activity in the forward pass (see Equation 3). The weight changes induced by each word were scaled by a logarithmic compression of its frequency of occurrence (Kuçera & Francis, 1967). In the limit of small weight changes, this is equivalent to using frequency to alter the number of times a word is presented during training (cf. Seidenberg & McClelland, 1989), and it enables the use of a procedure for adapting the learning rate of each connection independently (Jacobs, 1988). It also has the advantage of allowing the use of any range of frequencies (see Plaut et al., 1994, for simulations involving training with actual word frequencies). During each epoch of training, the scaled weight changes are accumulated for each word in the training corpus in turn, at which point the weights are actually changed and the process is repeated. Thus, the training procedure only approximates the more psychologically appropriate but less computationally tractable procedure of updating the weights after each word presentation.

Testing procedure. A major advantage of the current phonological representation over that used by Seidenberg and McClelland (1989) is that it is completely straightforward to determine the pronunciation the network gives to any letter string. The ordering of phonemes in Table 3 embodies the relevant phonotactic constraints, although the network is insensitive to this ordering. Accordingly, the response of the network to any orthographic input can be determined simply by activating the appropriate grapheme units, running the network as described earlier, and then scanning the phonemes in left-to-right order and concatenating all active phonemes. For consonants, this was all phonemes with activity above 0.5. Because each monosyllabic pronunciation must contain exactly one vowel, only the most active vowel phoneme was included in the response. If the units for an affricate and each of its component phonemes are active (e.g., /ps/, /s/, and /p/), then the order of the components in the response is reversed from their standard order (e.g., /ps/ rather than /sp/).

The Coltheart et al. (1993) Rules

Deriving the rules. Coltheart et al. (1993) derived 144 GPC rules from the 2,897 words used by Seidenberg and McClelland (1989). Examples of the rules are given in Table 4. The rules were derived by the following method. Words were presented in random order. For each word, the algorithm inferred the GPC rules that describe the relationship between that word's orthography and phonology. The rules are stored in a rule base, and when a rule is derived, its frequency in the rule base is incremented by 1. Rules in this rule base carry with them an indication of the position of the letter within the word from which the rule was created: "b" for beginning, "e" for end, and "m" for between the beginning and end. Thus, the rules are position specific.

Single-letter rules are derived when the number of letters in the word equals the number of phonemes in the corresponding phonology. In this case, the algorithm assumes a simple one-to-one mapping (e.g., for MINT the rules are m →

Table 4
Examples of the Coltheart, Curtis, Atkins, and Haller (1993) Rules and Their Application

Word	Correct	Rule 1	Rule 2	Rule 3	Rule 4	Output
lame	lAm	l → l	a _ e → A	m → m		lAm
lamp	lamp	l → l	a → a	m → m	p → p	lamp
lance	lans	l → l	a _ e → A	n → n	ce → s	lAns[a]
land	land	l → l	a → a	n → n	d → d	land
lane	lAn	l → l	a _ e → A	n → n		lAn
lap	lap	l → l	a → a	p → p		lap
laps	laps	l → l	a → a	p → p	s → s	laps
lapse	laps	l → l	a _ e → A	p → p	se → s	lAps[a]
lard	lord	l → l	(CS)ar → o	r → r	d → d	lord
large	lorj	l → l	a _ e → A	r → r	ge → j	lArj[a]
sloe	slO	s → s	l → l	oe → O		slO
sloop	slUp	s → s	l → l	oo → U	p → p	slUp
slop	slop	s → s	l → l	o → o	p → p	slop
slope	slOp	s → s	l → l	o _ e → O	p → p	slOp
slot	slot	s → s	l → l	o → o	t → t	slot
slouch	slWC	s → s	l → l	ou → W	ch → C	slWC
slough	sl^f	s → s	l → l	ough → W		slW[a]
slow	slO	s → s	l → l	ow → W		slW[a]
sludge	sl^j	s → s	l → l	u → ^	dge → j	sl^j
slug	sl^g	s → s	l → l	u → ^	g → g	sl^g
sluice	slUs	s → s	l → l	ui _ e → I	ce → s	slIs[a]
slum	sl^m	s → s	l → l	u → ^	m → m	sl^m

Note. Key to pronunciations: A = a in DAY, a = a in BAT, O = o in BOAT, o = o in HOT, U = oo in GOOSE, ^ = u in BUT, e = ea in HEAD, I = i in DIVE, W = ow in HOW, cs = context-sensitive rule.
[a] = Error.

/m/, i → /i/, n → /n/, t → /t/). Multiletter rules are necessary when the word has more letters than phonemes (e.g., TACK has four letters but only three phonemes). These rules are derived by first applying the single-letter rules to the word and then using what is left over to create multiletter rules (e.g., ck → /k/). Letters in multiletter rules need not be adjacent, however (e.g., for GATE, rule a _ e → /A/ is derived). A special case of multiletter rules occurs when the word contains "silent" letters that do not contribute to the phonology of the word (e.g., the second F in BLUFF). These rules are formed by incorporating the letter immediately preceding the letter in question in the rule (e.g., ff → /f/). If the letter in question is at the beginning of the word, then the letter that follows it is used to form the rule (e.g., in KNIT, rule would be kn → /n/).

Some context sensitivity must be introduced because the algorithm otherwise generates conflicting rules (e.g., a → /a/ in HAM, but a → /o/ in HARM). If handled only by retaining the most frequent rule within a set of conflicting rules, the algorithm's ability to generate correct pronunciations would be severely compromised. Before discarding the less frequent rules, the absolute and relative frequencies of the rules are examined. If these frequencies are above 5 and 0.2, respectively, the algorithm tabulates the letters immediately preceding and following the letter in question. If a particular letter context "greatly predominates," it is used in the creation of a context-sensitive rule. In its current version, the algorithm only derives context-sensitive rules from words with the same number of letters and phonemes.

Position-dependent rules are consolidated into "a" type (all-position) rules when two of the three types of rules (b, e, and m) are represented in the rule list for a given GPC (e.g., the oo → /U/ rule is an "a" rule because "e" and "m" rules exist in the rule list before consolidation). This allows the application of the rule to novel occurrences (e.g., oo at the beginning of a word).

Applying the rules. The database of rules is then applied in generating pronunciations for letter strings (words and nonwords). Whether a rule is applied depends on a critical frequency parameter. The results that Coltheart et al. (1993) reported used a minimum frequency of 2: Rules had to be derived from at least two words in order to be used. When presented with a letter string, the algorithm applies the rules left to right, using the multiletter rules first and the single letter rules last. When a rule is used, the matching part of the input string is absorbed, and the rule application process is repeated for the remainder of the string.

For example, when presented with the word CHIP, the algorithm first checks for rules dealing with the entire string CHIP. Because it finds none, it then looks for rules dealing with CHI. After another unsuccessful search, it then looks for rules for CH and finds ch → /C/. The phoneme /C/ is stored, and the remainder of the input string, IP is then checked. Because no rule exists for IP, I is checked, and a single-letter rule is found (i → /i/). The phoneme /i/ is then stored. The string P is then presented for rule application, and the single-letter rule p → /p/ is found, causing the phoneme /p/ to be appended to the output. The output is thus /Cip/, which is correct.

The settings of the parameters in the Coltheart et al. (1993) rule generation and application algorithms yielded highly accurate performance on nonwords and errors on about 22% of the words in the corpus. The latter items are the exceptions to the rules, which must be listed separately. Table 5 shows a summary the kinds of items that are mispronounced. The most common error is a regularization of an irregularly pronounced word (e.g., ARE → /Ar/, DONE → /dOn/. The rules also fail to pick up on many subregularities. For example, ow is always pronounced as in HOW, causing mispronunciations of KNOW, BLOW, FLOW, GLOW, and so on. The other types of errors occur much less frequently. In assessing this model's performance on the nonword corpus, we used the same parameter settings as in the simulations reported in Coltheart et al.

Table 5
Types and Examples of Mispronunciations Produced by the Coltheart, Curtis, Atkins, and Haller (1993) Rules

1. Regularizations of obvious exception words

Item	Regularization
are	Ar
done	dOn
doubt	dWbt
pint	pint
tongue	tongU

2. Failures to pick up on subregularities

Pattern	Pronunciation	Causes errors on
ow	always W	flow, blow, glow, low, etc.
initial c	always k	cent, cease, cell, cite, etc.
initial g	always g	gene, gent, germ, gel, etc.
initial ph	always ph	initial ph → f

3. Errors on words with highly regular neighborhoods

Item	Pronunciation	Causes errors on
car	kar	car, bar, far, par, etc.
all	al	ball, fall, tall, mall, etc.
halt	h*t	halt, malt, salt
mind	mind	mind, find, rind, kind, etc.
cook	kUk	book, look, hook, took, etc.

4. Overgeneralization of rules

Item	Pronunciation	Overgeneralized rule
farce	fArs	a _ e → A
glance	glAns	a _ e → A
volt	vOt	ol → O
buck	bk	bu → b[a]

5. Other errors

Item	Pronunciation
path	pT
faith	fTi
shoal	SO
sixth	sikT

Note. The pronunciations are written with the symbols used in "A Distributed Developmental Model of Word Recognition and Naming" by M. S. Seidenberg and J. L. McClelland, 1989, *Psychological Review, 96,* p. 533. Copyright 1989 by the American Psychological Association. See the Appendix for a pronunciation key.
[a] Vowel is deleted from all bu_ words (e.g., bug → /bg/, but → /bt/, etc.).

Scoring Issues

Nonwords do not have conventional pronunciations, which introduces a question as to which responses should be scored as correct for the purpose of comparing models. Our guiding principle was that consistent criteria be used in scoring the models' output and the subjects' responses. Two analyses were performed. The liberal scoring criterion approximated the ones used in the Glushko (1979) and McCann and Besner (1987) studies: The models' responses to these items were scored as correct if we could identify a plausible basis for them (either a rule or an analogy to a neighboring word). For the nonwords in our study, we used the following criterion: As noted previously, the two most common pronunciations of each nonword accounted for more than 93% of subjects' responses. We scored the models' output for a nonword as correct if it matched either of these pronunciations. Thus, all three sets of behavioral data and both models were scored using approximately the same criteria.

For the data collected in our experiment, a stricter criterion was also used: We examined the distribution of pronunciations across subjects and determined how often each model produced the first, second, or third most common pronunciation. For example, there were two pronunciations for BLEA: /blE/, produced by 70.8% of the subjects, and /blA/, produced by 20.8% (the other responses were clear errors). The three pronunciations for FEANT were, in order of frequency, /fEnt/, /fent/, and /fAnt/. We determined how often the models' response for a given nonword matched one of these responses. The stricter criterion could not be used for Glushko's (1979) and McCann and Besner's (1987) items because these authors did not provide information about the alternative pronunciations generated by the subjects. In summary, the lenient criterion indicates the extent to which the models were producing plausible output for three sets of nonwords; the stricter criterion indicates the extent to which they matched subjects' preferences for the nonwords in the larger corpus.

For the PDP model, we used the weights that were used in simulations described by Plaut and McClelland (1993). For the rules, we used the 144 rules that were used in the simulations described by Coltheart et al. (1993).

Results

After 3,200 epochs of training, the network correctly pronounced all but 10 of the words in the training corpus (99.7% correct). Coltheart et al. (1993) reported that their model produced correct responses for about 78% of the words in Seidenberg and McClelland's (1989) 2,897 word corpus. The 22% that were mispronounced are considered exceptions, to be pronounced by a separate mechanism.

Turning to nonwords, we first consider performance on Glushko's (1979) and McCann and Besner's (1987) benchmark lists, using the lenient criterion. Glushko's nonwords are relatively simple items such as BINT and HEAN; McCann and Besner's are harder items such as JINJE and TUNCE. Both models performed comparably to subjects and differed little from each other (see Table 6). The results were similar for the 590 items in our study, except that performance of the PDP model was somewhat more accurate and closer to the subjects' than were the pronunciation rules.

Results using the stricter scoring criterion are presented in Figure 3. The figure indicates that the models matched subjects' preferences about equally well. The PDP model matched subjects' most preferred pronunciations 3.2% more often that did the rules. The other major difference was in the "other" category, in which the Coltheart et al. (1993) rules generated more pronunciations that were produced by subjects with low frequency or not at all.

These results are relevant to issues raised by Coltheart et al. (1993, p. 603). They examined the pronunciation of two nonwords, NIND and JOOK, in detail. Their rules generated the pronunciations /nind/ (rhymes with SINNED) and /jUk/ (rhymes with SPOOK), which were the pronunciations that subjects preferred. The Seidenberg and McClelland (1989) model produced the pronunciations /nInd/ (like KIND) and /juk/ (like BOOK). The PDP model erred on these items because it used information concerning the word bodies contained in these items. The most common pronunciation of -IND is as in FIND, MIND, and BIND; the most common pronunciation of -OOK is as in COOK, BOOK, and LOOK. The analyses of these two items suggested to Coltheart et al. that the PDP approach would be at a disadvantage in pronouncing nonwords because it is sensitive only to word bodies. In fact, the model is not restricted to information about word bodies; this unit is merely the most salient one. More important, the data presented in Figure 3 suggest that when a broader range of nonwords is considered, both models generate less preferred pronunciations about equally often. Thus, although the modified PDP model still pronounced JOOK as the less preferred /juk/, the rules pronounced SEART as the less preferred /sErt/, BRILD as the less preferred /brild/, and JEALM as the less preferred /jElm/.[6]

Finally, we considered the consistency effects in Table 1. The Coltheart et al. (1993) model does not make specific latency predictions for words or nonwords. The model assumes that nonwords are pronounced by applying the rules; hence, a factor such as the number of alternative pronunciations associated with a nonword should not be relevant, contrary to the results in Table 1. Thus, in its present state, the Coltheart et al. model does not account for these effects. We derived latency predictions from the PDP model as follows. The Seidenberg and McClelland (1989) model computed pronunciations in a single step, and

[6] These examples illustrate how the limitations of one's current model provide insights that point toward future developments. One reason why people avoid the pronunciation /juk/ is probably that whereas there are many words containing /uk/, there are none containing /ju/ (note that the vowel in JUDGE and JUG is different). The absence of this consonant–vowel combination derives from articulatory constraints that are outside the scope of the current model. Incorporating these constraints would be a natural direction for future research.

Table 6
Percentage Correct on Nonword Pronunciation, Lenient Scoring Criteria

Experiment	Subjects	PM	CCAH
Glushko (1979)	94.9	96.5	98.0
McCann & Besner (1987)	91.5	85.6	88.2
The current study	92.7	88.3	82.0

Note. Subject data are from the original experiments. PM = Plaut and McClelland's (1993) model; CCAH = output from Coltheart, Curtis, Atkins, and Haller's (1993) pronunciation rules.

naming latencies were simulated using a sum of squared error score. The revised model computes phonological codes over a series of time steps, and a closer analog of reaction time is provided by the number of steps for the output pattern to settle (i.e., for the activations of units to stop changing). The timescale of the simulation is determined by the τ parameter in Equations 3–4. For this analysis, the criterion for settling was that no unit state changes by more than 0.001 (i.e., a very small amount) and τ was set to 0.01 (i.e., 100 unit updates per unit of time). Table 1 provides summary data concerning nonwords associated with one, two, or three pronunciations across subjects. The model's mean settling times for these items are given in the last column of Table 1. The model's output is computed deterministically; hence, only one pronunciation is produced for each nonword. The settling times for these pronunciations show a distinct consistency effect: They increase as a function of the number of associated pronunciations, as in the subject data. A one-way analysis of variance on these data yielded a significant effect of type, $F(2, 539) = 23.516$, $p < .001$, with post hoc comparisons yielding highly significant differences between all pairs of means.

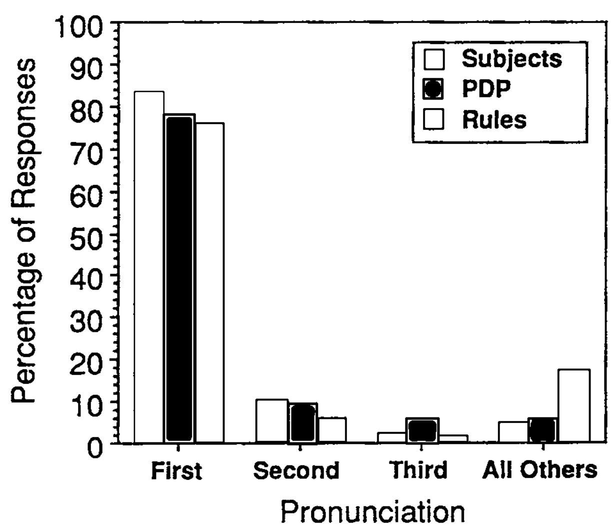

Figure 3. Fit between subjects' pronunciation preferences and models' responses. Data for subjects indicate percentage of responses accounted for by first, second, and third most common pronunciations and all others. Data for the models indicate the percentage of responses that matched subjects' first, second, third, or other pronunciations. PDP = parallel distributed processing model.

As expected, then, with the modifications introduced by Plaut and McClelland (1993) and the same scoring criteria applied to both model and people, the PDP model no longer exhibits a deficit in nonword performance compared with the rules. Moreover, it simulates the consistency effect without requiring additional stipulations.[7]

General Discussion

We have presented the results of a nonword naming experiment that provides data against which current models can be assessed. Subjects showed a high degree of agreement about the pronunciations of simple nonwords; however, many nonwords did generate alternative pronunciations that need to be considered. Both models produce plausible nonword pronunciations. The dual-route model does so by using rules that generate correct output for most words but mispronounce a significant number of exceptions. The PDP model does so by finding a set of weights that also allow it to generate correct output for more than 99% of the words, including both "rule-governed" items and "exceptions." Thus, simple PDP networks can encode both types of items with good generalization, as Seidenberg and McClelland (1989) suggested. Conjectures about the limitations of these networks by Besner et al. (1990), Pinker (Pinker & Prince, 1988; Prasada & Pinker, 1993), and Coltheart et al. (1993) are not supported by these results.

Plaut and McClelland (1993) carried out a number of analyses aimed at clarifying how the network is capable of reading both exception words and nonwords as well as skilled readers can. The one most relevant here involved trying to determine whether the network had segregated itself over the course of training into the two functionally separate subsystems of the dual-route model. If this were true, some hidden units should be particularly important for pronouncing nonwords but not exception words, whereas others should show the opposite specialization. Plaut and McClelland (1993) considered a hidden unit "important" for pronouncing an input if the cross-entropy error increased by at least 0.025 when that unit was removed from the network. The specific value of this criterion is not critical; the value of 0.025 was chosen so that approximately 20% of hidden units were considered important for a given stimulus on average. The segregation hypothesis suggests there should be a negative correlation across hidden units in the number of nonwords versus the number of exception words for which each is important. By contrast, for a set of orthographically matched nonwords and exception words (Tara-

[7] We have not conducted statistical tests to examine whether the small differences in performance seen in Figure 3 are reliable because the data do not justify this kind of comparison. We have made no attempt to vary parameters that would yield slightly different sets of rules or weights that might have small effects on the data. The results are sufficient to show that neither model is disadvantaged in terms of nonword performance.

ban & McClelland, 1987), there was a moderate positive correlation ($r = .43, p < .001$). Some hidden units are more important than others overall, but there is no evidence that the network has segregated itself into separate rule-based and lexical lookup mechanisms.

Having established that both models produce plausible nonword pronunciations, it is necessary to consider how they account for more detailed aspects of human performance. Our behavioral study replicated the consistency effect discovered by Glushko (1979): Nonwords containing spelling patterns associated with multiple pronunciations yielded longer latencies than nonwords that were assigned a single pronunciation. This effect also occurs for words (e.g., MINT takes longer to read than MUST because -INT is also pronounced as in PINT). These effects emerge naturally in PDP models because a single mechanism is used in reading both regular and irregular words; the weights therefore reflect exposure to both types of items and encode the degree of consistency in the mapping between spelling and sound. We have shown that the model described here also exhibits the consistency effect for nonwords.

These effects are problematical for the dual-route model; in fact, their discovery was taken as strong evidence against the assertion that regular words and nonwords are named by applying GPC rules (Glushko, 1979; Henderson, 1982; Patterson & Coltheart, 1987). Once the consistency effects were uncovered, it was necessary to introduce new assumptions into the dual-route model in order to account for them. These additional assumptions introduce new problems, however. Consistency effects are said to arise when the two routes yield conflicting information (Coltheart et al., 1993). Consider, for example, the irregular word PINT, the rule-governed but inconsistent word MINT, and the nonword BINT. PINT takes longer to name than an entirely regular word (e.g., BENT); in the dual-route model this is attributed to a conflict between the output of the lexical route (which is the correct pronunciation of PINT) and the output of the GPCs (the regularized pronunciation /pint/). Entirely regular, rule-governed words such as BENT do not produce this conflict; hence, they are named faster. This account does not extend gracefully to inconsistent words such as MINT, which also take longer to name than entirely regular and consistent words because of interference from exceptions such as PINT (e.g., Glushko, 1979; Jared, McRae, & Seidenberg, 1990). In the dual-route model, it must be assumed that processing of the word MINT results in activation of the pronunciation of PINT by means of the lexical route. To the extent that the lexical route is activating such "neighboring" words, it begins to approximate the analogy process described by Glushko (1979; see Patterson & Coltheart, 1987, for a discussion) as well as the effects of neighboring words on the weights in the PDP models. Thus, the dual-route model accommodates these effects by implementing mechanisms analogous to those in the PDP models.[8]

The consistency effect also occurs in nonwords (e.g., BINT is named more slowly than BIST); hence, it has to be assumed that attempting to name BINT also results in activation of the pronunciation associated with PINT. In the dual-route model this pronunciation can be accessed only through the lexical route. Thus, the model accounts for consistency effects by assuming that MINT and BINT activate the pronunciation of PINT the same way that PINT does. This entails abandoning the core assumption that nonwords are pronounced through the exclusive use of nonlexical GPCs, without any recourse to lexical knowledge.

In the PDP models, these effects follow from independently established principles about distributed representations and error-correcting learning algorithms. Moreover, these principles also account for frequency effects, the interaction of frequency and consistency, and a variety of other phenomena (Seidenberg & McClelland, 1989). The two models thus represent highly different approaches to explaining the behavioral phenomena. The PDP approach shows that various behavioral phenomena in reading follow from basic properties of learning in certain types of networks. The issues that this approach face concern things such as finding general solutions to the problem of representing phonological information, issues that are not at all specific to reading. The dual-route model starts with mechanisms that were introduced largely in response to broad patterns of impairment associated with acquired forms of dyslexia (e.g., Patterson, Coltheart, & Marshall, 1985). The issues that face this approach concern the validity of additional assumptions that need to be introduced in order to account for more detailed aspects of reading performance, such as consistency and frequency effects. The dual-route approach is therefore much more in the spirit of fitting models to data rather than deriving models from more general explanatory principles (Seidenberg, 1993). In closing, we consider the issues that confront each approach in a bit more detail.

Dual-Route Model

Coltheart et al. (1993) succeeded in generating a relatively small set of rules that yield good nonword performance. This is a significant advance over previous dual-route models, which relied heavily on the concept of GPCs without specifying their content. The specific rules discussed by Coltheart et al. introduce some problems that could be addressed by modifying the rules; however, they also raise more general issues concerning the validity of the approach.

The basic question regarding the proposed rules is whether they are the ones that people actually know and use in generating word and nonword pronunciations. Several issues arise. First, many of the rules lack face validity. For example, the algorithm induces the rule sh _ k → S because of the words SHACK, SHOCK, and SHUCK. This generates errors on words such as SHIRK, SHARK, and SHANK (the final /k/ is omitted), which then must be treated as exceptions. There is no independent evidence that people generate such unusual rules or that these items are exceptions. In other

[8] Coltheart, Curtis, Atkins, and Haller (1993) proposed incorporating an entire connectionist network (McClelland & Rumelhart, 1981) to serve as the lexical route.

cases, the rules produce correct output but for apparently spurious reasons. For example, BOSS is correctly pronounced by applying three rules:

b → /b/

o _ s → /*/

ss → /s/.

The second rule, which converts the pattern o _ s to the vowel /*/, is problematical. The rule is created by the LOSS–CROSS–BOSS–GLOSS neighborhood. However, it causes errors on words such as DOTS and GOES, which must be treated as exceptions. Moreover, it produces bizarre errors on these items (e.g., DOTS → /d*t/, GOES → /g*e/).

A third question is whether the rules correctly differentiate the rule-governed items from the exceptions, as they are supposed to do according to the dual-route theory. The number of exceptions that the rules are allowed to miss is not determined by independent evidence about people's performance on these words. The purpose of the rule-induction algorithm is to induce a set of rules that produces highly accurate performance on regular words and nonwords; the exceptions are then whichever words the rules fail to pronounce correctly. This set includes words that are generally agreed on to be exceptions (e.g., ACHE, ARE) and words that are not (e.g., BALL, DOTS). The rule and PDP models also make different predictions about which words should be difficult to pronounce. The rules, for example, treat SPOOK (the only -OOK word pronounced with /U/) as rule governed and COOK, BOOK, LOOK, TOOK, ROOK, HOOK, NOOK, BROOK, CROOK, and SHOOK as exceptions. This is because there is one rule governing OO, and it assigns the pronunciation that occurs in words such as FOOD, LOOP, and SOON. In the PDP models, the pronunciation of a vowel is affected by the context in which it occurs, particularly the coda. Thus, OO is pronounced /u/ when followed by K but /U/ when followed by N or P. According to the rules, SPOOK is easy and SHOOK is hard; for the PDP models, the opposite is predicted. Such predictions can be tested in behavioral experiments with normal subjects.

More generally, what these examples reveal is that the GPC-based approach cannot encode sub- or partial regularities, despite their prominence in English. Thus, the rules fail to encode the subregularity concerning the effects of coda K on nucleus OO. Moreover, this approach treats COOK, BOOK, and all of their neighbors as unrelated. They are simply items for which the rule for OO fails to generate correct output; the fact that COOK and BOOK also rhyme, owing to a generalization concerning the -OOK neighborhood, is completely missed. Insofar as such generalizations have a systematic impact on human naming performance (see Jared et al., 1990, for a summary), this is a problem for the GPC-based approach. One possibility would be to abandon the idea that the rules operate over graphemes and phonemes in favor of a more flexible system that operates over different-sized units. That would represent a return to a view developed by Shallice, Warrington, and McCarthy (1983). The other alternative is to abandon the commitment to the rule formalism entirely in favor of a type of representation that is better suited to capturing "quasi-regular" (Seidenberg & McClelland, 1989) forms of knowledge such as English spelling–sound correspondences. That is what the PDP models provide.

The rules also make predictions about the behavior of surface dyslexic patients that seem problematical in light of existing evidence. According to the dual-route theory, surface dyslexics have partial impairment in the "lexical" naming mechanism; thus, they pronounce exception words by applying rules, producing regularization errors. Patient MP (Bub, Cancelliere, & Kertesz, 1985) is a particularly notable case because almost all of her errors seemed to be clear regularizations. Thus, she pronounced STEAK as /stEk/, HAVE as /hAv/, and so on. The Coltheart et al. (1993) rules provide a basis for predicting which exception words should be regularized by such patients, namely, the items on which the rules produce regularization errors. In many cases the proposed rules make correct predictions: For example, they regularize STEAK and HAVE. However, in many cases they do not. For example, the rules generate correct pronunciations for FOUGHT, HYMN, POLL, SUIT, and TOMB, all of which were mispronounced by Patient MP (e.g., TOMB pronounced /tOm/ and SUIT pronounced /sUit/).

The question as to whether the rules isolate the correct set of exceptions has implications that transcend the particular rules proposed by Coltheart et al. (1993). The basic requirement for their algorithm is that it induce rules that generate correct output for nonwords and for regular words. The number of exception words that can be missed is relatively unconstrained. This introduces an important extra degree of freedom in the theory. As Seidenberg (1992a) noted, any system can be treated as rule governed if there is a second mechanism for dealing with all of the exceptions to the rules and no limit on what can be counted as an exception. Thus, the past tense in English can be treated as rule governed if one excludes exceptions such as SING–SANG and RING–RANG, and the spelling–sound correspondences of English can be treated as rule governed if enough words are treated as exceptions. As long as attention focuses only on producing the correct output, the dual-route model cannot fail: There are two types of phenomena (rule-governed cases and exceptions) and two mechanisms (rules, lexical lookup). As the discussion of Patient MP's data suggested, however, which items are treated as exceptions is actually an empirical question that needs to be addressed in order to eliminate this extra degree of freedom. A stronger assessment of the adequacy of the dual-route approach can be achieved by considering a broader range of phenomena than just pronunciation accuracy, eliminating the extra degree of freedom that currently exists. These phenomena include pronunciation latencies for different types of words and nonwords (see Coltheart and Rastle, 1994), as well as patterns observed in normal reading acquisition and developmental dyslexia.

Finally, there is a question as to how much the modified dual-route model differs from the PDP account. The Coltheart version of the dual-route model initially made two strong assumptions: (a) Regular words and nonwords are pronounced by applying nonlexical rules and (b) the rules operate over graphemes and phonemes without regard to context (see, e.g., Coltheart, 1978). The more recent dual-route model abandons both of these assumptions in favor of alternatives that make it more difficult to distinguish from the PDP approach. With regard to the first assumption, the naming of at least some regular words and nonwords must be assumed to involve both routes in order to accommodate the consistency effects. With regard to the second assumption, Coltheart et al. (1993) introduced a degree of context sensitivity into their rules, which means that they no longer simply refer to graphemes and phonemes. Context sensitivity increases the descriptive power of the rules enormously. However, it forfeits the appealing simplicity of the GPC idea and makes the rules' behavior hard to differentiate from that of the PDP network.[9]

The fact that rules are generated for correspondences that occur in as few as two items also contributes to this tendency. The number of items in which a GPC is found before it is added to the set of rules is a parameter, which was set to 2 in the Coltheart et al. (1993) study. This means that the "lexical" route is responsible for correspondences that only apply once. As we have shown, the models are already difficult to tell apart. Setting this parameter to 1, however, results in a model in which all words and nonwords can be pronounced by means of a single mechanism, exactly as in the PDP approach. A similar outcome can be achieved by exploiting the context sensitivity that Coltheart et al. allow in their rules. Once the rules are allowed to be context sensitive, there is nothing to prevent creating rules that generate correct pronunciations for exception words. With some additional assumptions (e.g., "strengths" associated with individual rules; conflicts between the rules as the source of consistency effects), the "rule-based" mechanism might succeed in simulating the behavior of the PDP network. At that point one could say that the rules provide an alternative means of implementing the net (see Seidenberg, 1992a, in press, for a discussion). This might be a useful thing to do because it would contribute to identifying the deeper underlying principles that govern behavior in this domain, but it would mean that the dual-route model does not provide a distinct theoretical alternative.

Of course, some of the resemblance between the PDP and dual-route models derives from the fact that both are dealing with the same phenomena. For example, both models must have mechanisms for computing from orthography to semantics and from semantics to phonology. This necessarily implies the existence of a second source of information relevant to generating pronunciations from print (see Plaut & Shallice, 1993, for models of this process). That this component of the lexical system contributes to the pronunciation of some words is supported by both computational and empirical evidence. The Seidenberg and McClelland (1989) model never learned the pronunciations of a small number of low-frequency irregular words (e.g., AISLE). Such items are good candidates for pronunciation by means of the orthography → semantics → phonology computation. Strain, Patterson, and Seidenberg (in press) provide evidence that semantic information is used in generating the pronunciations of such words. More generally, Plaut et al. (1994) discussed issues concerning what they termed the *division of labor* between components of the lexical processing system (see also Seidenberg, 1992b). Although this approach does not retain the idea of independent processing "routes" or other assumptions of the dual-route model, it is a response to many of the same issues that motivated the earlier approach.

PDP Models

The PDP model described here represents part of a series of experiments using different architectures to solve the orthography–phonology mapping problem (see Plaut et al., 1994). Such experiments provide important information about the factors that control a model's performance. Plaut and Shallice (1993), for example, examined a broad range of architectures relevant to the computation of word meanings and were able to identify general principles that gave rise to target phenomena. Our exploration of alternative architectures for performing the orthography to phonology mapping suggests that properties of the orthographic and phonological representations exert considerable influence over detailed aspects of performance. This was suggested by Seidenberg and McClelland's (1990) analysis of their model's errors, and it is borne out by the simulations discussed here.

Issues concerning the design of such representations are considered by Plaut et al. (1994). Plaut and McClelland's (1993) solution to the dispersion problem they identified in the Wickelphonology representation used by Rumelhart and McClelland (1986) and Seidenberg and McClelland (1989) involved two major changes: (a) replacing the phonetic feature triples of the Wickelphonology with a phonological level of representation and (b) introducing a syllabic structure (onset, nucleus, and coda positions) that allowed the representation to encode more of the phonotactic constraints of English. The new representation incorporates some additional aspects of phonological structure and eliminates unattractive features of the Wickelphonology that caused spurious errors. It should be clear, however, that the system described in Table 3 is only a further step toward completely general solutions to the problem of representing orthographic and phonological information. The representations we have used are still limited to monosyllables, and extensions to multisyllabic words are nontrivial. Moreover, future research will have to address how these orthographic and phonological representations develop. In reality, phonological representations are determined by constraints on possible segments imposed by articulatory and perceptual capacities and by characteristics of the language to which the child is exposed. Complex representations of the sound patterns of a language are in place before the child begins to read. These representations may themselves change as a

[9] The context-sensitive grapheme–phoneme correspondences might be termed *Wickelrules* (M. C. MacDonald, personal communication, 1993).

consequence of exposure to written language, especially alphabetic orthographies (Bertelson & de Gelder, 1989). The PDP models that we have described obviously did not attempt to model these developmental events. Rather, the revised model built some of the relevant orthotactic and phonotactic knowledge into the representations, which allowed us to focus on the problem of learning the mapping between them while maintaining good generalization. The better performance of the revised model is consistent with the view that having a highly structured phonological representation in place facilitates the acquisition of reading skill. Harm, Altmann, and Seidenberg (1994) describe simulations providing additional support for this conclusion. Their simulations showed that learning the correspondences between orthography and a prestructured phonological representation produced faster learning, more accurate asymptotic performance, and better nonword generalization than when the phonological representation was unstructured. These simulations point to directions for future research that will yield even more realistic models than the ones discussed here.

References

Bechtel, W., & Abrahamsen, A. (1991). *Connectionism and the mind.* Cambridge, MA: MIT Press.

Berko, J. (1958). The child's learning of English morphology. *Word, 14,* 150–177.

Bertelson, P., & de Gelder, B. (1989). Learning about reading from illiterates. In A. M. Galaburda (Ed.), *From reading to neurons* (pp. 1–23). Cambridge, MA: MIT Press.

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

Bub, D., Cancelliere, A., & Kertesz, A. (1985). Whole-word and analytic translation of spelling to sound in a non-semantic reading. In K. E. Patterson, J. C. Marshall, & M. Coltheart (Eds.), *Surface dyslexia: Neuropsychological and cognitive studies of phonological reading* (pp. 15–34). Hillsdale, NJ: Erlbaum.

Castles, V., & Coltheart, M. (1993). Varieties of developmental dyslexia. *Cognition, 47,* 149–180.

Coltheart, M. (1978). Lexical access in simple reading tasks. In G. Underwood (Ed.), *Strategies of information processing* (pp. 151–215). San Diego, CA: Academic Press.

Coltheart, M. (1987). Functional architecture of the language-processing system. In M. Coltheart, G. Sartori, & R. Job (Eds.), *Cognitive neuropsychology of language* (pp. 1–26). Hillsdale, NJ: Erlbaum.

Coltheart, M., Curtis, B., Atkins, P., & Haller, M. (1993). Models of reading aloud: Dual-route and parallel distributed processing approaches. *Psychological Review, 100,* 589–608.

Coltheart, M., & Rastle, K. (1994). Serial processing in reading aloud: Evidence for dual-route models of reading. *Journal of Experimental Psychology: Human Perception and Performance, 20,* 1197–1211.

Dell, G. (1986). A spreading activation theory of retrieval in sentence production. *Psychological Review, 93,* 283–321.

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

Harm, M., Altmann, L., & Seidenberg, M. S. (1994). Using connectionist networks to examine the role of prior constraints in human learning. In *Proceedings of the Sixteenth Annual Conference of the Cognitive Science Society* (pp. 392–396). Hillsdale, NJ: Erlbaum.

Henderson, L. (1982). *Orthography and word recognition in reading.* San Diego, CA: Academic Press.

Hinton, G. (1989). Connectionist learning procedures. *Artificial Intelligence, 40,* 185–234.

Humphreys, G. W., & Evett, L. J., (1985). Are there independent lexical and nonlexical routes in word processing? An evaluation of the dual-route theory of reading. *Behavioral and Brain Sciences, 8,* 689–740.

Jacobs, R. (1988). Increased rates of convergence through learning rate adaptation. *Neural Networks, 1,* 295–307.

Jared, D., McRae, K., & Seidenberg, M. S. (1990). The basis of consistency effects in word naming. *Journal of Memory and Language, 29,* 687–715.

Jared, D., & Seidenberg, M. S. (1991). Does word recognition proceed from spelling to sound to meaning? *Journal of Experimental Psychology: General, 120,* 358–394.

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

Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *Annals of Mathematical Statistics, 22,* 79–86.

McCann, R. S., & 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.

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, 86,* 287–330.

Patterson, K. E., & Coltheart, V. (1987). Phonological processes in reading: A tutorial review. In M. Coltheart (Ed.), *Attention and performance XII: Reading* (pp. 421–448). Hillsdale, NJ: Erlbaum.

Patterson, K. E., Marshall, J. C., & Coltheart, M. (1985). *Surface dyslexia: Neuropsychological and cognitive studies of phonological reading.* Hillsdale, NJ: Erlbaum.

Pearlmutter, B. (1989). Learning state space trajectories in recurrent neural networks. *Neural Computation, 1,* 263–269.

Pinker, S. (1991). Rules of language. *Science, 253,* 530–534.

Pinker, S., & Prince, A. (1988). On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition, 28,* 73–194.

Plaut, D. C., & McClelland, J. L. (1993). Generalization with componential attractors: Word and nonword reading in an attractor network. In *Proceedings of the Fifteenth Annual Conference of the Cognitive Science Society,* (pp. 824–829). Hillsdale, NJ: Erlbaum.

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. E. (1994). *Understanding normal and impaired word reading: Computational principles in quasi-regular domains.* Manuscript submitted for publication.

Plaut, D. C., & Shallice, T. (1993). Deep dyslexia: A case study of connectionist neuropsychology. *Cognitive Neuropsychology, 10,* 377–500.

Prasada, S., & Pinker, S. (1993). Generalization of regular and irregular morphological patterns. *Language and Cognitive Processes, 8,* 1–56.

Rumelhart, D. E., Durbin, R., Golden, R., & Chauvin, Y. (in press). Backpropagation: The basic theory. In D. E. Rumelhart & Y. Chauvin (Eds.), *Backpropagation: Theory and practice.* Cambridge, MA: MIT Press.

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

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: Explorations in the microstructure of cognition* (Vol. 2, pp. 216–271). Cambridge, MA: MIT Press.

Seidenberg, M. S. (1989). Word recognition and naming: A computational model and its implications. In W. D. Marlsen-Wilson (Ed.), *Lexical representation and process* (pp. 52–69). Cambridge, MA: MIT Press.

Seidenberg, M. S. (1992a). Connectionism without tears. In S. Davis (Ed.), *Connectionism: Advances in theory and practice* (pp. 84–122). New York: Oxford University Press.

Seidenberg, M. S. (1992b). Beyond orthographic depth: Equitable division of labor. In R. Frost & L. Katz (Eds.), *Orthography, phonology, morphology, and meaning* (pp. 85–118). Amsterdam: North-Holland.

Seidenberg, M. S. (1993). Connectionist models and cognitive theory. *Psychological Science, 4,* 228–235.

Seidenberg, M. S. (in press). Visual word recognition: An overview. In P. Eimas & J. L. Miller (Eds.), *Handbook of perception and cognition: Language.* San Diego, CA: Academic 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., & McClelland, J. L. (1990). More words but still no lexicon: Reply to Besner et al. *Psychological Review, 97,* 447–452.

Sejnowski, T. J., & Rosenberg, C. R. (1987). Parallel networks that learn to pronounce English text. *Complex Systems, 1,* 145–168.

Selkirk, E. (1982). The syllable. In H. van der Hulst & N. Smith (Eds.), *The structure of phonological representations* (pp. 337–383). Dordrecht, The Netherlands: Foris.

Shallice, T., Warrington, E. K., & McCarthy, R. (1983). Reading without semantics. *Quarterly Journal of Experimental Psychology, 35A,* 111–138.

Strain, E., Patterson, K. E., & Seidenberg, M. S. (in press). Semantic contributions to word naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition.*

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

Treiman, R. (1992). The role of intrasyllabic units in learning to read and spell. In P. Gough, L. Ehri, & R. Treiman (Ed.), *Reading acquisition* (pp. 65–106). Hillsdale, NJ: Erlbaum.

Van Orden, G., Pennington, B. F., & Stone, G. O. (1990). Word identification in reading and the promise of a subsymbolic psycholinguistics. *Psychological Review, 97,* 488–522.

Venezky, R. (1970). *The structure of English orthography.* The Hague, The Netherlands: Mouton.

Williams, R. J., & Peng, J. (1990). An efficient gradient-based algorithm for on-line training of recurrent network trajectories. *Neural Computation, 2,* 490–501.

(Appendix follows on next page)

Appendix

Data Concerning Preferred Pronunciations of 590 Nonwords

Item	Pron	*n*	RT	Item	Pron	n	RT
baint	bAnt	23	671	clart	klort	24	654
bange	banj	14	732	cleash	klES	23	793
barce	bors	24	669	clert	klert	23	615
barsh	borS	23	691	cles	kles	22	628
bartz	borts	21	654	cleve	klEv	20	770
baugh	b*	19	712	clise	klIz	10	715
beese	bEs	15	634	clo	klO	23	597
beil	bAl	11	718	cloor	klUr	17	630
beint	bAnt	18	694	clurt	klert	23	641
belf	belf	24	658	clyle	klIl	17	916
belm	belm	23	655	cooze	kUz	22	652
bense	bens	22	666	cound	kWnd	20	729
bibe	bIb	24	612	craid	krAd	24	644
bierce	bErs	21	766	crame	krAm	20	665
biff	bif	24	623	creal	krEl	24	640
bimpse	bimps	23	686	crean	krEn	23	731
binc	bink	21	655	creet	krEt	24	689
binch	binC	23	607	creighth	krAT	14	899
bint	bint	24	624	crelt	krelt	24	658
bip	bip	24	650	crent	krent	24	651
bipe	bIp	22	658	cryke	krIk	22	685
blaft	blaft	22	635	cuce	kUs	21	827
blan	blan	24	574	curnt	kernt	23	748
blash	blaS	24	615	dacht	dakt	13	658
blaunt	blont	20	663	dade	dAd	24	604
blea	blE	17	654	dafe	dAf	18	642
blex	bleks	23	614	dain	dAn	24	62[illegible]
blypt	blipt	14	743	dar	dor	24	613
boach	bOC	24	638	dask	dask	24	579
boaf	bOf	21	696	daste	dAst	17	692
boarse	bOrs	24	719	dath	daT	24	592
boist	bYst	22	628	datt	dat	24	620
bonge	bonj	13	770	dench	denC	24	607
borge	bOrj	21	642	denth	denT	24	570
bort	bOrt	24	594	derch	derC	21	663
bove	bOv	19	681	dewt	dUt	21	632
braist	brAst	24	677	diend	dEnd	10	860
braze	brAz	23	631	dieve	dEv	15	703
breat	brEt	18	737	dilge	dilj	23	650
brewn	brUn	23	694	dilt	dilt	24	576
brild	brild	23	722	dirm	derm	23	627
brist	brist	24	609	disp	disp	24	575
brune	brUn	24	692	dithe	dID	8	675
buch	b^C	11	754	dixth	diksT	20	667
bup	b^p	24	642	doath	dOT	22	618
burf	berf	24	632	dode	dOd	22	640
bux	b^ks	22	625	doir	dOr	12	724
byth	biT	9	944	dold	dOld	17	612
cack	kak	23	669	doof	dUf	22	635
chamb	Cam	18	769	doup	dUp	21	675
chank	Cank	21	681	draille	drAl	22	723
chape	CAp	19	751	drang	draN	23	635
chaut	C*t	12	711	drase	drAs	12	602
chawn	C*n	21	706	dre	drE	14	645
chazz	Caz	21	673	drebb	dreb	24	665
chence	Cens	21	795	dreer	drEr	23	680
cherd	Cerd	21	662	drel	drel	24	620
chig	Cig	23	658	drept	drept	21	631
chis	Ciz	13	793	drit	drit	20	677
choad	COd	22	723	droap	drOp	21	692
choll	COl	13	670	drock	drok	24	620
chone	COn	20	722	dront	dront	22	656
chork	Cork	22	690	drook	drUk	19	669
chure	Cer	16	757	drost	drost	21	648
chye	CI	22	706	drouth	drWT	18	673

Appendix (continued)

Item	Pron	*n*	RT	Item	Pron	n	RT
drow	drW	16	619	glab	glab	24	653
druile	drUl	17	797	glarc	glork	20	737
drust	dr^st	22	609	glay	glA	23	625
duess	dUs	14	797	glealth	glET	12	714
duge	dUj	16	753	gleard	glErd	19	711
duilt	dilt	10	794	glebt	glebt	17	696
dur	der	17	665	glep	glep	21	659
durb	derb	24	649	glesh	gleS	20	660
durse	ders	24	665	glief	glEf	18	762
dush	d^S	24	687	glithe	gliT	9	711
dyst	dist	17	714	glourt	glOrt	13	728
	faS/			gluff	gl^f	23	641
fache	faC	11	794	glusk	gl^sk	23	682
fane	fAn	23	706	goak	gOk	24	742
fard	ford	23	663	goise	gYz	17	699
faunch	f*nC	22	665	golk	g*lk	14	643
fauze	f*z	14	706	gomb	gom	17	703
feant	fEnt	14	785	gou	gU	18	696
feap	fEp	23	733	grall	grol	14	656
feath	fET	18	739	graw	gr*	23	643
feech	fEC	19	611	graxe	graks	15	653
fey	fA	21	655	greep	grEp	24	677
fich	fiC	21	626	greft	greft	23	646
fiek	fEk	13	656	grend	grend	24	644
finth	finT	22	673	grood	grUd	23	645
fipt	fipt	21	724	groot	grUt	24	627
firch	ferC	24	699	gruite	grUt	15	806
firk	ferk	23	679	gruy	grU	10	657
fiss	fis	23	654	gulb	g^lb	22	662
fize	fIz	13	804	habe	hAb	24	604
flas	flas	22	634	halm	halm	12	688
fleik	flAk	10	790	hapt	hapt	19	666
floth	floT	19	665	heam	hEm	22	727
flun	fl^n	23	700	hease	hEs	11	744
flutch	fl^C	22	693	hef	hef	24	583
foast	fOst	20	698	hegg	heg	24	681
fod	fod	24	619	hene	hEn	19	744
fonce	fons	23	642	hength	heNT	19	780
fong	f*N	23	621	hepth	hepT	21	671
fooch	fUC	18	641	herf	herf	23	658
foon	fUn	23	668	herge	herj	22	643
foose	fUs	21	717	hifth	hifT	16	761
forch	fOrC	23	640	hile	hIl	20	600
foun	fWn	12	780	hilk	hilk	23	633
foupe	fUp	22	726	hine	hIn	23	676
frad	frad	21	678	hink	hink	24	636
frand	frand	20	771	hisk	hisk	23	583
frast	frast	20	698	hoat	hOt	22	668
freamt	frEmt	17	865	hodd	hod	23	632
frell	frel	22	645	hoost	hUst	22	651
fren	fren	23	678	hoothe	hUD	15	609
frewe	frU	21	707	horst	horst	24	595
froke	frOk	24	664	hosh	hoS	20	698
fru	frU	24	674	howd	hWd	22	712
fruise	frUz	12	816	huild	hUld	13	968
fruke	frUk	23	719	hulp	h^lp	23	632
fruse	frUz	20	713	jamp	jamp	23	649
fulse	f^ls	21	651	jate	jAt	24	631
fumb	f^m	15	667	jauce	j*s	13	718
furk	ferk	24	651	jealm	jelm	17	811
garge	gorj	23	626	jeir	jEr	17	747
gat	gat	23	596	jide	jId	23	626
gerse	gers	21	679	jind	jind	23	631
gick	gik	22	690	jir	jer	20	631
gieze	gEz	11	847	jitt	jit	24	629
gign	gin	10	839	jom	jom	22	664

(Appendix continues on next page)

Appendix (continued)

Item	Pron	*n*	RT	Item	Pron	n	RT
jope	jOp	23	598	nerr	ner	17	626
jore	jOr	24	718	nerth	nerT	22	648
jum	j^m	23	676	nid	nid	23	612
jyre	jIr	21	779	niest	nEst	12	789
kag	kag	23	598	nilth	nilT	21	739
kail	kAl	22	670	ninx	ninks	23	709
karch	korC	21	613	nire	nIr	13	708
keace	kEs	18	766	noil	nYl	22	694
kearn	kern	17	706	nooth	nUT	19	719
kess	kes	23	686	norld	nOrld	21	716
kie	kI	21	664	nounge	nWnj	13	745
kirth	kerT	23	633	nouse	nWs	20	689
kith	kiT	20	663	nowth	nWT	17	835
konze	konz	21	711	nube	nUb	21	647
koust	kWst	17	740	nuck	n^k	22	696
kurst	kerst	23	712	nurge	nerj	22	657
larb	lorb	23	581	padge	paj	24	594
larf	lorf	24	568	paff	paf	24	559
leige	lEj	18	710	pait	pAt	24	591
lext	lekst	22	680	palk	p*k	14	621
lirge	lerj	22	680	pawk	p*k	23	616
litz	lits	24	610	ped	ped	24	615
loice	lYs	24	659	pelve	pelv	24	638
lorce	lOrs	24	629	peme	pEm	19	700
lorm	lOrm	24	635	pern	pern	23	616
lourse	lOrs	17	643	peud	pyUd	14	768
lourth	lerT	8	725	pice	pIs	19	653
ludge	l^j	23	645	pidst	pidst	20	730
luzz	l^z	20	616	piege	pEj	11	831
mact	makt	20	640	pift	pift	24	666
mage	mAj	19	713	pight	pIt	21	751
malve	malv	14	712	plaive	plAv	23	685
manch	manC	20	700	plap	plap	24	611
mangst	maNst	17	719	plear	plEr	24	643
marn	morn	22	712	plerk	plerk	22	603
masp	masp	23	624	plew	plU	24	660
meave	mEv	22	708	plewd	plUd	22	647
meeve	mEv	22	689	pling	pliN	23	686
meird	mErd	18	718	plon	plon	24	613
meize	mEz	11	681	ploop	plUp	23	633
melch	melC	24	627	plourn	plOrn	11	662
melfth	melfT	20	738	plown	plWn	16	689
mier	mEr	15	660	plox	ploks	24	581
mim	mim	23	714	plue	plU	23	644
mird	merd	23	668	plut	pl^t	21	605
mirst	merst	23	679	pog	p*g	22	631
mish	miS	23	623	poin	pYn	24	635
miz	miz	20	726	poove	pUv	24	643
moax	mOks	23	643	poss	pos	23	667
modge	moj	22	739	pral	prol	12	646
moft	m*ft	23	653	preadth	predT	18	725
moid	mYd	23	665	preel	prEl	23	676
molf	m*lf	22	653	prot	prot	24	605
motch	moC	22	610	pude	pUd	15	647
moung	mWN	8	748	puit	pwEt	9	921
mource	mOrs	18	751	pult	p^lt	22	706
mourd	mOrd	16	748	pung	p^N	22	647
moy	mY	24	641	purn	pern	24	609
muest	mUst	16	748	pute	pyUt	16	630
mulge	m^lj	21	659	pymn	pim	20	723
mulk	m^lk	20	627	raim	rAm	22	621
murd	merd	24	637	ralp	ralp	19	647
myp	mip	16	764	rance	rans	24	629
nadd	nad	24	591	rause	rWs	8	646
naise	nAz	19	645	rawl	r*l	22	626
narve	norv	23	672	reast	rEst	23	721
nault	n*lt	21	662	redge	rej	23	606
neak	nEk	24	702	reeze	rEz	24	623
neld	neld	22	675	relte	relt	24	658

Appendix (continued)

Item	Pron	*n*	RT	Item	Pron	n	RT
rem	rem	24	562	spauge	sp*j	9	857
ret	ret	24	548	spaul	sp*l	21	749
ri	rI	12	676	speight	spAt	12	943
rield	rEld	19	703	spetch	speC	23	805
riew	ryU	12	937	spoan	spOn	24	713
rike	rIk	24	592	spowl	spWl	18	764
ril	ril	23	651	spram	spram	20	728
ringe	rinj	23	635	staltz	st*lts	19	823
rive	rIv	22	623	stamn	stam	22	749
rix	riks	22	573	starp	storp	24	715
ronk	r*nk	23	597	steach	stEC	20	784
roo	rU	24	609	stearth	sterT	17	830
roosh	rUS	24	620	stimp	stimp	24	760
rould	rUld	7	723	stoze	stOz	21	685
rounce	rWns	21	657	strop	strop	20	820
rud	r^d	20	637	sturch	sterC	24	726
rull	r^l	21	593	sule	sUl	16	827
rund	r^nd	23	630	sump	s^mp	24	710
rynch	rinC	21	761	sunge	s^nj	15	754
saisle	sAl	7	911	surl	serl	24	667
sanse	sans	17	686	sutt	s^t	23	739
scole	skOl	24	740	sym	sim	20	692
scra	scr*	18	792	sype	sIp	22	725
seaf	sEf	22	753	tace	tAs	23	642
seart	sert	16	823	tald	t*ld	20	636
seb	seb	21	689	tark	tork	24	581
sempt	sempt	23	730	tarmth	tOrmT	24	736
sheapt	SEpt	11	927	tarse	tors	24	624
shearse	SErs	15	842	tatch	taC	24	666
shelk	Selk	20	745	tays	tAz	23	660
sial	sIl	16	757	tearch	terC	14	720
sib	sib	22	687	tearl	terl	16	692
sidth	sidT	23	784	teigh	tA	9	856
silm	silm	20	708	teign	tAn	16	820
simb	sim	17	760	telp	telp	24	643
skose	skOz	10	825	terb	terb	23	578
slere	slEr	15	741	terve	terv	24	620
slote	slOt	20	782	thak	Tak	24	687
smair	smAr	24	770	thealt	Telt	13	796
smalse	sm*ls	19	812	thoar	TOr	24	682
smapse	smaps	21	892	thout	DWt	22	741
smaught	sm*t	19	842	thrax	Traks	24	691
smead	smEd	22	729	thwee	TwE	19	828
smein	smAn	10	890	tidge	tij	20	676
smill	smil	23	766	tiece	tEs	11	785
smough	smW	9	792	tinse	tins	23	662
smoul	smWl	9	746	tirl	terl	24	658
smuard	smord	13	793	tirt	tert	22	671
smuice	smUs	20	952	titch	tiC	21	675
smuide	smUd	13	858	toal	tOl	23	599
snass	snas	20	752	toard	tOrd	23	615
snauve	sn*v	14	815	tob	tob	24	577
sneed	snEd	20	713	tolve	tOlv	17	646
sneue	snU	21	820	tord	tOrd	24	591
snoam	snOm	21	830	torl	tOrl	22	598
snoud	snWd	12	730	torse	tOrs	24	624
snurr	sner	21	737	toubt	tWt	13	833
sny	snI	22	849	traph	traf	19	710
sobe	sOb	23	701	trave	trAv	19	644
somp	somp	24	690	treathe	trED	12	734
sond	sond	21	716	treek	trEk	22	653
sorn	sOrn	24	648	treen	trEn	23	691
sorth	sOrT	19	694	trest	trest	22	661
sount	sWnt	16	674	troes	trOz	22	753
spake	spAk	24	766	trome	trOm	22	682
spathe	spAD	10	727	trool	trUl	23	672

(Appendix continues on next page)

Appendix (continued)

Item	Pron	*n*	RT	Item	Pron	n	RT
troom	trUm	23	657	welse	wels	20	681
trouch	trWC	20	649	wese	wEz	11	755
trunt	tr^nt	24	630	wesk	wesk	22	661
tuede	tUd	8	740	wict	wikt	20	732
tulf	tulf	20	648	wolt	wOlt	17	617
tunch	t^nC	22	632	wompt	wompt	22	585
tusp	t^sp	22	650	worpse	wOrps	15	673
tymph	timf	21	770	wouge	wUj	12	682
vaight	vAt	20	911	wounge	wWnj	14	724
valf	v*lf	9	697	woute	wUt	14	784
vant	vant	18	640	wug	w^g	23	614
vaud	v*d	14	702	wulch	w^lC	20	634
vect	vekt	23	624	wunk	w^nk	22	646
veef	vEf	20	726	wurve	werve	21	684
ver	ver	15	634	yalt	y*lt	24	636
vime	vIm	21	734	yare	yAr	20	634
vinn	vin	23	661	yarm	yorm	23	621
voe	vO	23	684	yied	yEd	11	723
voint	vYnt	23	737	yife	yIf	19	672
vor	vOr	24	615	yin	yin	23	650
vought	v*t	9	746	yince	yins	24	673
vour	vUr	8	746	yoops	yUps	21	659
vub	v^b	23	704	yothe	yOT	7	628
vyme	vIm	21	709	yus	y^s	20	670
waith	wAT	23	661	zale	zAl	24	712
weck	wek	23	653	zigh	zI	12	754
weem	wEm	21	665	zisle	zIl	9	776
weeth	wET	22	671	zuct	z^kt	19	779
weg	weg	22	627	zuss	z^s	18	772

Note. Pron = most common pronunciation; *n* = number of subjects out of 24 providing that pronunciation; RT = reaction time (naming latency in milliseconds). Pronunciation key: A = a as in PLAY; a = a as in BAT; E = ee as in BEET; e = e as in BET; O = o as in HOPE; o = o as in HOT; I = i as in BITE; i = i as in BIT; U = oo as in BOOT; u = oo as in BOOK; W = ow as in HOW; Y = oy as in BOY; ^ = u as in BUT; * = aw as in PAW; S = sh as in SHOE; C = ch as in CHEW; T = th as in THIN; D = th as in THIS. This set of symbols is adapted from "A Distributed Developmental Model of Word Recognition and Naming" by M. S. Seidenberg and J. L. McClelland, 1989, *Psychological Review, 96,* p. 533. Copyright 1989 by the American Psychological Association.

Received January 31, 1994
Revision received April 1, 1994
Accepted April 15, 1994 ■