Psychological Review
1982, Vol. 89, No. 5, 573-594

0033-295X/82/8905-0573$00.75

An Activation-Verification Model for Letter and Word Recognition: The Word-Superiority Effect

Kenneth R. Paap, Sandra L. Newsome, James E. McDonald, and Roger W. Schvaneveldt
New Mexico State University

An activation-verification model for letter and word recognition yielded predictions of two-alternative forced-choice performance for 864 individual stimuli that were either words, orthographically regular nonwords, or orthographically irregular nonwords. The encoding algorithm (programmed in APL) uses empirically determined confusion matrices to activate units in both an alphabetum and a lexicon. In general, predicted performance is enhanced when decisions are based on lexical information, because activity in the lexicon tends to constrain the identity of test letters more than the activity in the alphabetum. Thus, the model predicts large advantages of words over irregular nonwords, and smaller advantages of words over regular nonwords. The predicted differences are close to those obtained in a number of experiments and clearly demonstrate that the effects of manipulating lexicality and orthography can be predicted on the basis of lexical constraint alone. Furthermore, within each class (word, regular nonword, irregular nonword) there are significant correlations between the simulated and obtained performance on individual items. Our activation-verification model is contrasted with McClelland and Rumelhart's (1981) interactive activation model.

The goal of the activation-verification model is to account for the effects of prior and concurrent context on word and letter recognition in a variety of experimental paradigms (McDonald, 1980; Paap & Newsome, Note 1, Note 2; Paap, Newsome, & McDonald, Note 3; Schvaneveldt & McDonald, Note 4). An interactive activation model, inspired by the same set of sweeping goals, has recently been described by McClelland and Rumelhart (1981). Although the models complement one another nicely with regard to some aspects, we will contrast the two approaches in our final discussion and highlight the very important differences between them.

The verification model was originally developed to account for reaction time data from lexical-decision and naming tasks (Becker, 1976, 1980; Becker & Killion, 1977; McDonald, 1980; Schvaneveldt, & McDonald, 1981; Schvaneveldt, Meyer, & Becker, 1976; Becker, Schvaneveldt, & Gomez, Note 5). Although the various discussions of the verification model differ about certain details, there has been general agreement about the basic structure of the model. The basic operations involved in word and letter recognition are encoding, verification, and decision. We refer to the model described in the present paper as the activation-verification model to emphasize the extensive treatment given to encoding processes that are based on activation of letter and word detectors. The activation process shares many features with the logogen model proposed by Morton (1969). In the activation-verification model, we have attempted to formalize earlier verbal state-

Portions of this research were presented at the meetings of the Psychonomic Society, St. Louis, November 1980; the Southwestern Psychological Association, Houston, April 1981; and the Psychonomic Society, Philadelphia, November 1981. The project was partially supported by Minigrant Award 1-2-02190 from the Arts and Sciences Research Center at New Mexico State University. We would like to thank Ron Noel, Jerry Sue Thompson, and Wayne Whitemore for their contributions to various stages of this research. Also, we appreciate the thoughtful reviews of a first draft of this paper provided by Jay McClelland, Dom Massaro, and Garvin Chastain.

Sandra Newsome is now at Rensselaer Polytechnic Institute in Troy, New York. James McDonald is now at IBM in Boulder, Colorado.

Requests for reprints should be sent to Kenneth R. Paap, Department of Psychology, Box 3452, New Mexico State University, Las Cruces, New Mexico, 88003.

ments about the verification model. As we will show, this formalization permits a quantitative evaluation of aspects of the model with data from the word-superiority paradigm.

The activation–verification model consists of encoding, verification, and decision operations. Encoding is used to describe the early operations that lead to the unconscious activation of learned units in memory. In the case of words, the most highly activated lexical entries are referred to as the set of candidate words.

Verification follows encoding and usually leads to the conscious recognition of a single lexical entry from the set of candidates. Verification should be viewed as an independent, top-down analysis of the stimulus that is guided by a stored representation of a word. Verification determines whether a refined perceptual representation of the stimulus word is sufficiently similar to a particular word, supported by the evidence of an earlier, less refined analysis of the stimulus. This general definition of verification is sufficient for the current tests of the activation–verification model, but more specific assumptions have been suggested (e.g., Becker, 1980; McDonald, 1980; Schvaneveldt & McDonald, 1981) and could be the focus of future work. For example, verification has been described as a comparison between a prototypical representation of a candidate word and a holistic representation of the test stimulus. However, within the framework of our model, we could just as easily suggest that verification involves a comparison between the letter information available in an activated word unit and the updated activity of the letter units in the alphabetum.

The verification process has been instantiated in a computer simulation that mimics the real-time processing involved in verification (McDonald, 1980). The simulated verification process is a serial-comparison operation on the set of candidate words generated during encoding. Thus, verification results in a match or mismatch. If the degree of fit between the visual evidence and the candidate word exceeds a decision criterion, then the word is consciously recognized. If the match does not exceed the criterion, then the candidate is rejected and the next candidate is verified. Semantic context affects the definition of the candidate set, whereas word frequency affects the order of verification for words in the candidate set. Those words in the candidate set that are related to the context will be verified before those that are not. If the verification process finds no match among the set of related words, it proceeds to check the remaining candidates in a decreasing order of word frequency. These provisions produce semantic-priming and word-frequency effects in a simulated lexical-decision task. The upper panel of Figure 1 depicts the important structures and processes that are simulated for a typical lexical-decision task that involves normal stimulus durations of 250 msec or more.

The factors affecting the speed and accuracy of performance in a particular paradigm depend on whether decisions are based primarily on information from encoding or from verification. Because verification relies on a comparison that involves continuing perceptual analysis of the stimulus, the potential contribution of verification should be severely attenuated whenever a backward mask overwrites or erases the sensory buffer. Thus, paradigms that present masked letter strings offer a potential showcase for the predictive power of our simulated encoding process. The bottom panel of Figure 1 shows the reduced model that is appropriate for very short stimulus durations or stimuli that are masked.

Of primary importance is the model's ability to explain why letters embedded in words are recognized more accurately than letters embedded in nonwords. The current version of the model predicts not only this word-superiority effect (WSE) as a general phenomenon but also the relative performance for any given letter string. The predictions are derived from the following descriptions of the encoding process and the decision rule.

Encoding

Feature Matching

Like many others, we view encoding as a process that involves matching features to various types of units. The model assumes two types of units: whole words stored in a lexicon and individual letters stored in an

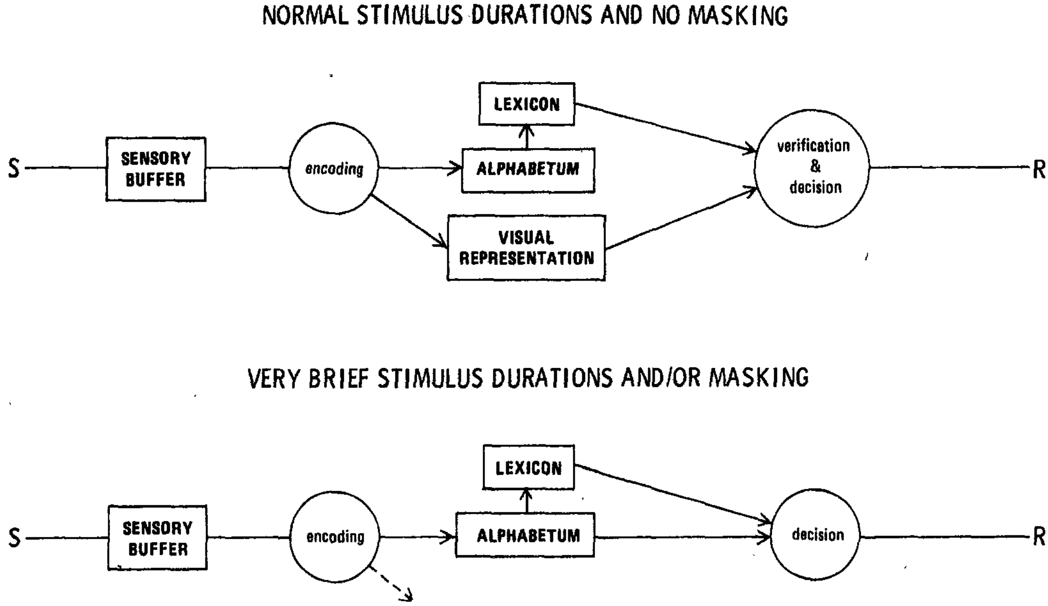

Figure 1. The upper panel shows the important structures that the model simulates for a typical lexical-decision task that involves normal stimulus durations of 250 msec or more; thé lower panel shows the reduced model that is appropriate for very short stimulus durations and/or stimuli that are masked.

alphabetum. Each letter of the alphabet is represented by a feature list, with the relative level of activation for each letter unit determined by the number of matching and mismatching features that have been detected. Word units are activated to the extent that their constituent letters are activated in the alphabetum. The model also allows for the possibility that the detection of supraletter features (e.g., word shape or word length) may directly contribute to the activation level of the word units. However, because the present evaluation of the encoding process consists entirely of four-letter uppercase strings, we have assumed that there are no distinctive supraletter features.

It is a straightforward matter to implement a simulation based on feature matching. However, this strategy is not likely to succeed because the selection of the appropriate set of features relies heavily on guesswork. If inappropriate features are used, a bogus set of candidate words will be generated.

Confusion Probabilities as Activation

To avoid the problem of selecting the correct set of features, the activation–verification model uses empirically determined confusion matrices to generate activation levels in the alphabetum and lexicon. Table 1 shows the obtained confusion matrix for the uppercase characters we used. Entries are the percentage of responses (columns) for each letter as a stimulus (rows). The specific procedure used to obtain this matrix has been reported elsewhere (Paap, Newsome, & McDonald, Note 3).

We assume that confusability reflects the degree of feature matching and the appropriate rules for combining matching and mismatching information. This definition of activation emphasizes the role of psychophysical distinctiveness because an identity match does not always lead to the same level of activation. For example, because the probabilities of a correct response given *K*, *S*, and *V* as stimuli (*K*/*K*, *S*/*S*, & *V*/*V*) are .748, .541, and .397, respectively, the model assumes that *S*, a letter of average confusability, receives less activation than the more distinctive letter *K*, but more activation than the less distinctive letter *V*.

All of the matrices used to generate predictions are transformations of the matrix shown in Table 1. Transformations are ap-

Table 1
Confusion Matrix for the Terak Uppercase Letters

Stimulus	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
A	45	6	0	1	2	2	2	8	0	1	2	2	0	2	1	1	1	16	2	1	1	0	0	1	1	1
B	3	61	0	2	4	2	3	2	1	1	2	3	0	1	2	2	1	4	2	0	2	1	1	1	0	0
C	1	1	54	5	3	1	3	1	0	1	1	2	0	2	9	1	2	3	3	1	2	0	0	1	0	1
D	1	0	0	66	1	0	1	2	2	0	3	3	1	2	8	1	1	0	3	0	2	0	1	0	0	1
E	1	4	0	1	65	6	2	3	0	1	2	3	0	1	0	1	0	3	4	0	0	0	1	0	0	1
F	1	2	0	1	11	64	1	2	1	1	1	1	0	0	1	2	0	3	2	1	0	0	1	0	0	0
G	2	3	2	4	1	2	61	1	1	0	2	2	0	1	4	0	3	1	1	2	1	0	0	0	1	1
H	2	2	0	1	0	1	2	73	0	1	2	1	1	1	1	1	0	2	1	1	1	0	1	1	1	1
I	1	1	0	1	1	4	2	5	53	2	2	6	0	3	1	1	1	2	1	6	2	1	1	1	1	1
J	1	0	1	4	1	1	3	2	6	41	2	4	0	2	4	0	0	2	1	4	11	2	2	2	0	1
K	1	0	1	1	1	1	3	2	0	0	75	2	1	3	0	0	0	1	0	1	1	0	1	2	0	0
L	1	2	1	2	1	0	0	2	2	1	2	64	1	1	2	0	0	2	1	2	5	1	2	1	0	1
M	1	0	0	1	2	1	2	6	2	2	3	2	56	10	0	1	0	2	1	1	0	0	2	2	1	0
N	0	0	0	2	3	0	1	3	1	1	2	1	1	76	1	1	0	1	0	0	2	0	1	1	0	0
O	1	1	2	10	2	0	3	2	1	0	1	1	0	1	58	1	6	1	1	1	2	0	0	1	0	0
P	3	3	0	0	3	2	4	2	1	1	2	2	0	1	1	60	0	9	1	1	1	0	0	1	1	1
Q	4	2	1	6	3	1	8	3	1	0	1	1	0	3	13	1	36	6	2	1	0	1	2	0	1	1
R	1	2	0	2	2	1	2	2	1	2	2	2	1	3	0	1	0	69	1	1	1	0	0	0	1	0
S	1	3	1	2	4	4	5	3	0	2	2	3	0	1	1	1	0	5	54	1	1	0	2	1	1	0
T	0	1	0	2	3	2	1	4	13	1	3	3	1	1	0	0	0	2	2	56	1	0	1	0	1	1
U	1	1	1	2	1	0	2	1	1	1	1	1	0	1	4	1	1	2	0	2	64	5	3	0	1	1
V	1	0	1	1	2	0	1	1	1	1	1	2	0	0	3	0	0	1	1	0	35	40	3	0	1	1
W	1	2	1	1	2	1	2	8	1	2	2	2	1	8	0	1	0	1	1	1	2	1	53	1	0	1
X	2	1	0	2	2	0	1	3	1	0	9	1	2	4	0	0	0	2	1	1	1	0	2	61	0	2
Y	1	1	0	1	1	1	1	6	0	1	2	2	1	3	2	0	0	2	1	2	5	1	3	1	57	1
Z	2	2	1	1	3	2	3	3	2	3	3	5	0	3	1	1	0	2	3	10	3	1	1	1	2	39

Note. Entries are the percentages of responses (columns) for each letter as a stimulus (rows).

plied to model any variable that is assumed to affect stimulus quality. For example, if the onset asynchrony between stimulus and mask is greater than the 17 msec used to generate the percentages shown in Table 1, then the values on the main diagonal (for correct responses) should be increased, whereas the off-diagonal values (for incorrect responses) are decreased. The particular adjustment used increases each correct response percentage by a percentage of the distance to the ceiling and decreases each incorrect response percentage by a percentage of the distance to the floor. The increments and decrements are such that the rows always sum to 100%. The procedure is reversed when stimulus quality is degraded rather than enhanced.

Another effect that the model can capture by appropriate transformations of the basic matrix is loss of acuity for letters at greater distances from the average fixation point. All of the predictions reported later access separate matrices for each of the four spatial positions. The extent to which separate matrices improve the model's predictions depends on whether correlations between obtained and predicted data are based on all stimulus items or only those that test the same target position. To demonstrate this we derived a single matrix in which each cell entry was the mean of the four confusion probabilities found in the separate matrices. When the single matrix is used, correlations between predicted and obtained performance are significantly higher for the subsets of stimuli that all share the same target position than across the entire set of stimuli. When separate confusion matrices are used, the correlation for the entire set of stimuli rises to about the same level as the separate correlations on each position.

As an example of how the encoding process uses the confusion matrices, consider the presentation of the input string PORE. As indicated in Figure 2, position-specific units in the alphabetum are assumed to be activated

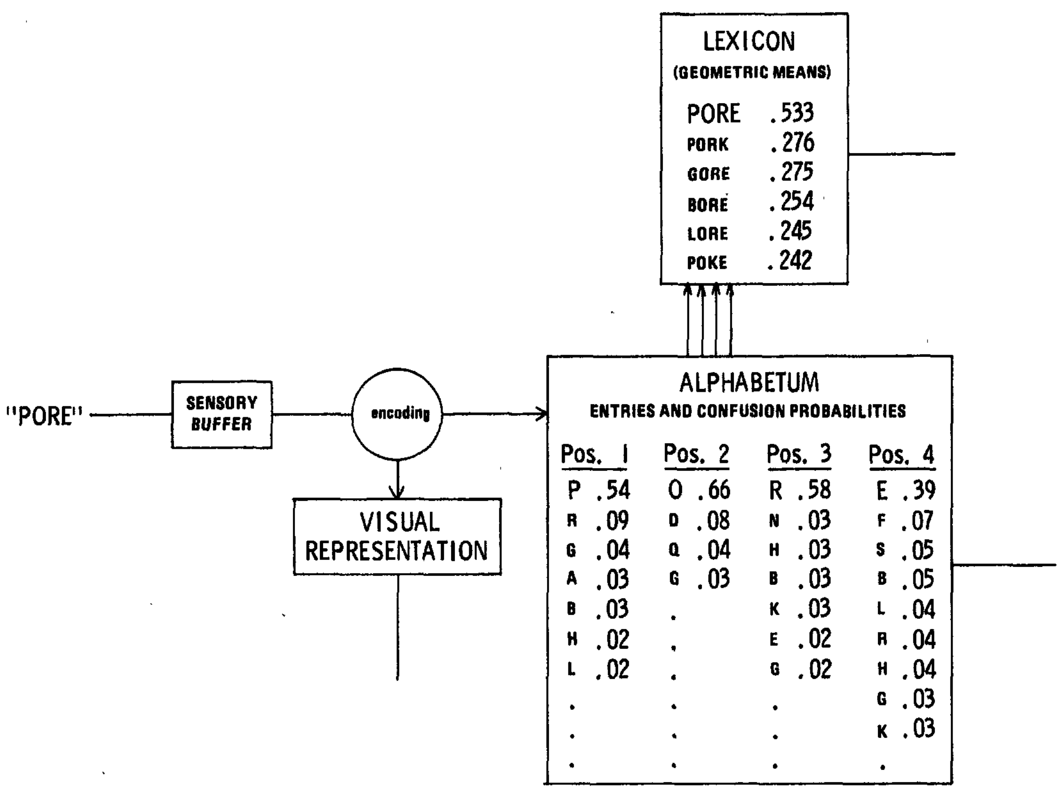

Figure 2. Encoding the word PORE. (Activation strengths for letter units in the alphabetum are determined by letter-confusion probabilities. Activation strengths for word units in the lexicon are determined by taking the geometric mean of the corresponding letter-confusion probabilities.)

in direct proportion to their confusability. In the first position the input letter *P* activates the corresponding *P* unit the most (.538), the *R* unit more than any other remaining unit (.091), and several other units (*G*, *A*, *B*, *H*, and *L*) to lesser extents. Patterns of activation are established in a similar manner for the other three spatial positions.

Activity in the alphabetum continuously feeds into the lexicon. The encoding algorithm estimates the activation strength for each word in the lexicon by taking the geometric mean of the activity levels associated with the constituent letters. One consequence of using the geometric mean is that one very inactive letter unit (close to zero) may prevent activation of a potential word unit that is receiving high levels of activation from three other letter units. This may mirror psychological reality because otherwise identical versions of the model yield poorer fits to the obtained data if the geometric mean is replaced by the arithmetic mean or the square root of the sum of squares (the vector distance between another word and the input word in a space generated from the letter-confusion probabilities).

The Word-Unit Criterion

The decision system does not monitor all of the activity in the lexicon. The model assumes that the activity in a word unit can be accessed by the decision system only if the level of activation exceeds a preset criterion. The predictions reported in this paper are all based on a word-unit criterion of .24. With this criterion word stimuli generate an average of about 3.4 words in the candidate set compared to about 2.1 words for stimuli that are orthographically regular pseudowords. If the word-unit criterion is raised, fewer words will be accessible to the decision system. In our final discussion we will suggest that a high criterion may offer an alternative explanation for the pseudoword-expectancy effect reported by Carr, Davidson, and Hawkins (1978).

For the example illustrated in Figure 2, six word units exceed the criterion for the input word PORE: PORE (.533), PORK (.276), GORE (.275), BORE (.254), LORE (.245), and POKE (.242). Nonwords can also activate the lexicon through the same mechanism. For example, when the pseudoword DORE is input to the simulation, three word units exceed a geometric mean of .240: DONE (.268), LORE (.265), and SORE (.261). Nonwords with lower levels of orthographic structure tend to produce less lexical activity. For example, when EPRO (an anagram of PORE) is presented to the encoding algorithm, no word units exceed the .240 criterion.

Decision

Decision Criterion

If the task requires detection or recognition of a letter from the stimulus, the decision process is assumed to have access to the relative activation levels of all units in the alphabetum and those units in the lexicon that exceed the word-unit criterion. It is further assumed that when total lexical activity exceeds some preset criterion, the decision will be based on lexical rather than alphabetic evidence. This decision criterion is different from the individual word-unit criterion, and the distinction should be kept clearly in mind. Exceeding a word-unit criterion makes that particular lexical entry accessible to the decision system. Exceeding the decision criterion leads to a decision based on lexical activity rather than alphabetic activity.

It is advantageous to base a decision on lexical evidence when there is some minimal amount of activation, because many words can be completely specified on the basis of fewer features than would be necessary to specify their constituent letters when presented in isolation. Accordingly, lexical candidates will tend toward greater veracity than alphabetic candidates whenever decisions are made on the basis of partial information.

The specific decision rules used to predict performance in a two-alternative, forced-choice letter-recognition task are as follows: For any stimulus, the predicted proportion correct (PPC) depends on contributions from both the lexicon and alphabetum. More specifically, PPC is the weighted sum of the probability of a correct response based on lexical evidence and the probability of a correct response based on alphabetic evidence:

$$\mathrm{PPC} = P(L) \times P(C/L) + P(A) \times P(C/A), \quad (1)$$

where $P(L)$ is the probability of a lexically based decision, $P(C/L)$ is the conditional probability of a correct response given that a decision is based on the lexicon, $P(A)$ is the probability of an alphabetically based decision, and $P(C/A)$ is the conditional probability of a correct response based on alphabetic information. Because the decision for each trial is made on the basis of either lexical or alphabetic information, $P(A)$ is equal to $1 - P(L)$.

Correct Responses From the Lexicon

The probability of a correct response given a decision based in the lexicon is

$$P(C/L) = 1.0 \times (\Sigma w_c/\Sigma w) + .5 \times (\Sigma w_n/\Sigma w) + 0 \times (\Sigma w_i/\Sigma w), \quad (2)$$

where Σw_c is the activation strength of word units that support the target letter, Σw_n is the activation strength of word units that support neither the correct nor the incorrect alternative, Σw_i is the activation strength of word units that support the incorrect alternative, and Σw is the total lexical activity.

The general expression for $P(C/L)$ shown in Equation 2 was selected for reasons of parsimony and programming efficiency. The equation can be viewed as the application of a simple high-threshold model (Luce, 1963) to each lexical entry. When a word unit exceeds the criterion, the decision system will (a) select the correct alternative with a probability of 1.0 whenever the letter in the critical position supports the correct alternative, (b) select the correct alternative with a probability of 0.0 whenever the letter in the critical position supports the incorrect alternative, and (c) guess whenever the critical letter supports neither alternative. The only additional assumption required is that the decision system combine the probabilities from each lexical entry by simply weighting them in proportion to their activation strengths. For the following examples, words had to exceed a criterion of .24 in order to be considered by the decision system.

If the decision for any single trial is based on lexical activity, our underlying process model assumes that something like Equation 2 does apply. That is, we have adopted the working hypothesis that decisions based on unverified lexical evidence involve a weighted strength of the word units supporting each of the two-choice alternatives. Alternatively, $P(C/L)$ could be viewed as the probability of certain word units being the most highly activated units on individual trials. We note as an aside that our general approach has been to find a set of simple algorithms (with plausible psychological underpinnings) that do a good job of predicting performance. An alternative approach is to begin with very specific ideas about the underlying psychological processes and then derive algorithms to suit these particular assumptions. We have shied away from this latter strategy in the belief that both the tests and selection of particular psychological explanations would be easier once we had developed a formal model that predicts performance in several paradigms with a fair amount of success.

The factors that determine the probability of a correct response from the lexicon can be easily understood by examining specific examples. If the stimulus word PORE is presented (see Figure 2) and the third position is probed with the alternatives *R* and *K*, we have

$$P(C/L) = 1 \times (1.583/1.825) + .5 \times (0/1.825) + 0 = .867. \quad (3)$$

This relatively high probability of a correct response is reasonable because five of the highly activated words (BORE, PORK, GORE, LORE, PORE) support the correct alternative, whereas only POKE supports the incorrect alternative. In general, $P(C/L)$ will be .70 or greater for words; but exceptions do occur. For example, when the word GONE is presented to the simulation, the following words, with their activation strengths in parentheses, exceed the cutoff: DONE (.281), GONE (.549), TONE (.243), BONE (.278), CONE (.256), and LONE (.251). If the first position is probed with the alternatives *G* and *B*, we have

$$P(C/L) = 1 \times (.549/1.858) + .5 \times (1.031/1.858) + 0 = .57 \quad (4)$$

Lower values of $P(C/L)$ tend to occur when there is a highly activated word that supports the incorrect alternative and/or when there are several highly activated words that support neither alternative.

Correct Responses From the Alphabetum

The probability of a correct response given a decision based on the alphabetum is

$$P(C/A) = 1.0 \times (a_c/\Sigma a) + .5 \times (\Sigma a_n/\Sigma a) + 0 \times (a_i/\Sigma a), \quad (5)$$

where a_c is the activation strength of the letter unit corresponding to the correct alternative, Σa_n is the activation strength of the letter units that are neither the correct nor the incorrect alternative, and Σa is the total alphabetic activity. The only difference between the decision rule for the alphabetum and that for the lexicon is that alphabetic activity is not filtered by a criterion.

Assuming that the third position is probed with the alternatives R and K, the $P(C/A)$ for the stimulus word PORE is

$$P(C/A) = 1 \times (.585/1.000) + .5 \times (.390/1.000) + 0 = .780. \quad (6)$$

This value would, of course, be the same for the pseudoword DORE, the anagram EPRO, or any other stimulus that contains R in the third position.

Probability of a Decision Based on the Lexicon

For any given trial, it is assumed that a decision will be made on the basis of lexical information if total lexical activity exceeds the decision criterion. Given noise introduced by variations in the subject's fixation or attention, and within the visual processing system itself, it is reasonable to assume that a specific stimulus will exceed or fall short of the decision criterion on a probabilistic, rather than an all-or-none, basis. Accordingly, the mathematical instantiation of our verbal model estimates, for each stimulus, the probability that its lexical activity will exceed the decision criterion. This probability will, of course, depend on both the average amount of lexical activity produced by the stimulus in question and the current value of the decision criterion.

The first step in estimating $P(L)$ normalizes the total lexical activity produced by each individual stimulus to that stimulus that produced the greatest amount of lexical activity. Of the 288 words that have been used as input to the encoding algorithm, the word SEAR has produced the greatest number of words above criterion (9) and the greatest amount of total lexical activity (2.779). Thus, normalization involves dividing the total lexical activity for a given stimulus by 2.779.

Normalization is simply a convenience to ensure that the amount of lexical activity generated by each stimulus will fall in the range of 0 to 1 and, consequently, that $P(L)$ will also be bounded by 0 and 1. Because this transformation simply involves dividing by a constant, we are not altering the relative lexical strengths that were initially obtained by summing the geometric means of all words above the word-unit criterion. In any event, we certainly do not mean to infer that subjects must somehow know in advance the greatest amount of lexical activity that they will experience during the course of the experiment. Rather, we simply assume that total lexical activity is one important determiner of $P(L)$.

The contribution of the decision rule to $P(L)$ is reflected by a second step that raises each of the normalized activation levels by a constant power between 0 and 1. This yields the estimated $P(L)$ for each stimulus. Stringent decision criteria can be modeled by using high exponents (near 1). This procedure generates a wide range of $P(L)$ across items, and a decrease in the average $P(L)$. Lax decision criteria can be modeled by using low exponents (near 0). A very lax criterion compresses the range toward the upper boundary and thus causes the mean $P(L)$ to approach 1. Consequently, when a very lax criterion is used, $P(L)$ tends to be quite high for any level of lexical activity. Using an exponential transformation is a convenient way to operationalize decision rules as diverse as "use lexical evidence whenever it is available" (exponents near 0) to "use lexical evidence only for those stimuli that produce substantial amounts of lexical activity" (exponents near 1). All of the predictions discussed later are based on a constant value (.5) for this parameter.

Because $P(L)$ is derived from total lexical activity, it will generally be the case that stimuli like PORE that excite six word units above threshold will have higher probabilities than stimuli like RAMP which produce only one suprathreshold word unit. In summary, the probability that a decision will be based on lexical evidence is estimated for each stimulus using the following equation:

$$P(L) = (W_i/W_{max})^n, \quad (7)$$

where W_i is the total lexical activity for stimulus i, W_{max} is the total lexical activity for the stimulus producing the greatest activity, and the exponent n is a parameter that reflects the stringency of the criterion. $P(L)$ for the stimulus PORE would be

$$P(L) = (1.825/2.779)^{.5} = .810. \quad (8)$$

When the exponent n is set to .5, $P(L)$ for word stimuli will range from about .4 to 1.0, with a mean of about .6.

Finally, it is assumed that when total lexical activity is less than the criterion, the decision will, by default, be based on alphabetic information. Accordingly, the probability of an alphabetic decision, $P(A)$, is

$$P(A) = 1 - P(L). \quad (9)$$

Predicted Probability Correct

Table 2 uses Equation 1 to show the derivation of the overall probability of a correct response for two sets of stimuli. Each set consists of a word, a pseudoword that shares three letters in common with the word, and an anagram of the word. The first set was chosen because it produces predictions that are similar to most sets of words and nonwords and illustrates why the model will yield different mean PPCs for words, pseudowords, and anagrams. The second set is abnormal and illustrates some principles that account for variations within stimulus classes.

As exemplified by PORE, the probability of a correct response based on lexical evidence is usually greater than that based on alphabetic evidence. The overall proportion correct falls somewhere between the lexical and alphabetic probabilities and will approach the lexical value as $P(L)$, the probability of a lexical decision, increases. In general, words should provide better context than nonwords to the extent that (a) $P(C/L) > P(C/A)$ and (b) $P(L)$ is high. Because these conditions are met for the stimulus PORE, the model predicts a 4.2% advantage over the pseudoword DORE and a 6.6% advantage over the anagram EPRO.

The model predicts that some words should actually produce word-inferiority effects. This can only occur, as in the example LEAF, when lexical evidence is poorer than alphabetic evidence. Because the probability of a lexical decision is estimated from total lexical activity, regardless of the veridicality of that information, the model predicts that LEAF will be judged on the basis of the inferior lexical evidence about two thirds of the time. This leads to a predicted 8.4% disadvantage relative to the pseudoword BEAF and a 6.1% disadvantage relative to the anagram ELAF.

Table 2
Simulation of Word, Pseudoword, and Anagram Differences for Two Examples

			Simulated values					
Class	Stimulus	Alternatives	WSE	SPC	$= P(L)$	$\times P(C/L)$	$+ P(A)$	$\times P(C/A)$
Typical								
Word	PORE	R, K		.852	= .810	× .867	+ .190	× .786
Pseudoword	DORE	R, K	+.042	.810	= .535	× .831	+ .465	× .786
Anagram	EPRO	R, K	+.066	.786	= .000	× .000	+ 1.000	× .786
Atypical								
Word	LEAF	F, P		.621	= .591	× .677	+ .323	× .682
Pseudoword	BEAF	F, P	−.084	.705	= .428	× .736	+ .572	× .682
Anagram	ELAF	F, P	−.061	.682	= .000	× .000	+ 1.000	× .682

Note. WSE = word-superiority effect; SPC is the simulated proportion correct; $P(C/L)$ is the probability of a correct response from the lexicon; $P(C/A)$ is the probability of a correct response from the alphabetum; and $P(L)$ is the probability of basing a decision on lexical information.

Test and Evaluation of the Model

The model can be tested at two levels. First, by averaging across stimuli in the same class, the model can be used to predict the magnitude of the WSE for words over pseudowords or words over anagrams. Second, the model should be able to predict item variation within a stimulus class.

Four experiments provide the basis for the following tests (Paap & Newsome, Note 1, Note 2; Paap, Newsome, McDonald, & Schvaneveldt, Note 6). All experiments used the two-alternative, forced-choice letter-recognition task. Each experiment compared performance on a set of 288 four-letter words to a set of 288 nonwords. The nonwords used in two of the experiments were orthographically regular pseudowords. In the remaining two experiments, the nonwords were formed by selecting that anagram for each word stimulus that minimized the amount of orthographic structure. The two alternatives selected for each stimulus both formed words for word stimuli and nonwords for the nonword stimuli.

Word and Pseudoword Advantages

Our first approach to evaluating the model was to use the algorithm described in the introduction to predict the proportion correct for each of the 288 words, pseudowords, and anagrams. The mean output of the model for words, pseudowords, and anagrams is shown in Table 3. The simulation predicts a 2.8% advantage for words (.841) over pseudowords (.813), and an 8.6% advantage for words over anagrams (.755). These differences compare favorably to the obtained WSEs of 2.6% and 8.8%, respectively.

Across all 288 words, the number of lexical entries exceeding the cutoff ranged from 1 to 9, with a mean of 3.4. These word units constrain the identity of the critical letter more effectively than it is constrained by the activity within the alphabetum. Thus, the word advantages predicted by the model occur because lexical information is used 63% of the time and the mean probability of a correct response from the lexicon (.897) is greater than that based on the alphabetum (.758).

The major reason why the model yields lower proportions correct for nonwords than words is not the quality of the available lexical evidence, but rather its frequent absence. That is, the probability of a correct response based on lexical evidence for the 253 pseudowords that produce at least one word above threshold is nearly identical (about .90) to that for the 288 words. Similarly, $P(C/L)$ for the 44 anagrams that produce at least one word above the cutoff is .94. Thus, the quantity and not the quality of lexical information is the basis for the WSE. Orthographically regular pseudowords excite the lexicon almost as much as words (2.1 vs. 3.4 entries) and lead to small word advantages, whereas orthographically irregular anagrams generate much less lexical activity (.2 vs. 3.4 entries) and show much larger word advantages.

Item-Specific Effects

The model's ability to predict performance on specific stimuli is limited by the sensitivity and reliability of the data. Our previous work provides two sets of word data and one set

Table 3
Simulated Values for Words, Pseudowords, and Anagrams

	Simulated values				
Lexical class	PPC	$P(C/L)$	$P(C/A)$	$P(L)$	NW
Words	.841	.897	.758	.634	3.4
Pseudowords	.813	.791	.758	.415	2.1
Anagrams	.755	.144	.758	.073	.2

Note. PPC is the predicted proportion correct; $P(C/L)$ is the probability of a correct response from the lexicon; $P(C/A)$ is the probability of a correct response from the alphabetum; $P(L)$ is the probability of basing a decision on lexical information; and NW is the number of words that exceeded the criterion.

for each of the two types of nonwords. Each of the 288 items in a set was presented to 24 different subjects. This means that the obtained proportions correct for individual items vary in steps of .04. Given these limitations, a correlation of data against data provides an index of the maximum amount of variation that could be accounted for by the model. The correlation between the two sets of word data was .56. A similar determination of the reliability of the pseudoword and anagram data yielded correlations of .48 and .39, respectively. However, because only 24 subjects saw each nonword stimulus, these lower correlations are due, in part, to the fact that each half consisted of only 12 observations compared with the 24 available in the word analysis.

Table 4 shows the correlations between the various sets of obtained data and the values generated by the model. Because each correlation is based on a large number (288) of pairs, significant values of r need only exceed .12. For all three stimulus classes, there are significant correlations between the obtained data and (a) the predicted proportion correct, (b) the probability of a correct response from the lexicon, and (c) the probability of a correct response from the alphabetum. The correlations are quite high considering the limitations discussed above. For example, the correlation between the first set of word data and the predicted proportion correct is .30 compared to .56 for data against data. Taking the ratio of the squared values of these correlations (.09 and .31, respectively) leads to the conclusion that the model can account for 29% of the consistent item variation (both correlations are based on 24 observations per data point, and no correction for n is needed).

As a final check on the model's ability to predict variation within words, the 288 words were partitioned into thirds on the basis of their predicted performance, and mean obtained performance was computed for each group. Obtained proportion correct for the upper third was .85 compared to .82, and .78 for the middle and bottom thirds.

The source of the model's success in predicting interitem variation is difficult to trace. Because decisions about word stimuli are made on the basis of lexical evidence more often than on alphabetic evidence, $P(L) =$ .63, it is clear that both the lexicon and alphabetum contribute substantially to the overall PPC, and accordingly, both branches must enjoy some predictive power in order to avoid diluting the overall correlation between obtained and predicted correct. Furthermore, it should be noted that the correlation between $P(C/L)$ and the obtained data is quite sensitive to the word-unit criterion (because this affects the average number of candidate words). This is consistent with the view that the predictive power of the lexical branch primarily depends on getting the correct set of candidate words and is not a simple transformation of alphabetic activity.

The item-specific predictions are far from exact, but they are quite encouraging because our lexicon contains only the 1,600 four-letter words listed in the Kucera and Francis (1967) norms. Because $P(C/L)$ for any item is determined by the activation strengths of visually similar words in the lexicon, substantial variation for a particular item can be

Table 4
Correlations Between Obtained Proportion Correct and Simulated Values

Stimulus type	Simulated values: PPC	$P(C/L)$	$P(C/A)$	$P(L)$	NW
Words					
Set 1	+.30	+.28	+.29	−.05	−.05
Set 2	+.26	+.23	+.27	+.01	.00
Anagrams	+.37	+.21	+.34	+.17	+.14
Pseudowords	+.35	+.17	+.38	+.15	+.16

Note. PPC is the predicted proportion correct; $P(C/L)$ is the probability of a correct response from the lexicon; $P(C/A)$ is the probability of a correct response from the alphabetum; $P(L)$ is the probability of basing a decision on lexical information; and NW is the number of words that exceeded the criterion.

introduced if just one highly similar word is either added or deleted from the lexicon.

Lexical Constraint

The test words consisted of the 288 words used by Johnston (1978) in his influential test of sophisticated-guessing theory. Half of the words were defined by Johnston as high-constraint words, and the other half as low-constraint words. Johnston assumed that lexical knowledge will constrain the identity of the critical letter in inverse proportion to the number of different letters that will form words given the remaining context. For example, the context _ATE supplies much less constraint than the context _RIP because 10 letters form words in the former context, but only three in the latter. Johnston rejected the hypothesis that lexical constraint contributes to the WSE because performance on the high-constraint words (.77) was slightly lower than performance on the low-constraint words (.80).

Our model shows that when the same partial information, in the form of letter-confusion probabilities, is provided to both the alphabetum and lexicon, lexical activity can support the critical letter more often than does the alphabetic activity. This difference between $P(C/L)$ and $P(C/A)$ provides an index of the potential amount of lexical benefit for any word. We view this measure of lexical benefit as an alternative definition for the global concept of lexical constraint. Thus, Johnston's (1978) conclusion that lexical constraint does not contribute to the WSE may have been premature and the product of a less appropriate definition of lexical constraint. Concerns that we have raised previously (Paap & Newsome, 1980a) can now be extended in the context of our model and the alternative definition for lexical constraint.

Johnston (1978) obtained both free-recall and forced-choice responses. First, consider those trials on which the three context letters were correctly reported. The conditional probabilities of a correct critical-letter report given a correct report of all three context letters were .90 and .86 for high- and low-constraint pairs, respectively. This is extremely high performance for free recall, and any significant differences due to lexical constraint may be obscured by a ceiling effect. Moreover, if one assumes that the same stimuli presented to the same subjects under the same conditions would yield performance distributions with some variability, then it would seem quite reasonable to characterize these trials as samples that have been drawn from the upper end of the distribution and that reflect trials on which the level of visual information was unusually high.

When stimulus information is high, the effects of lexical constraint may be low. Our model makes exactly this prediction. If stimulus quality is enhanced by transforming the correct responses in the confusion matrices upward, and the incorrect responses downward, the difference between the lexical and alphabetic branches disappear. For example, if stimulus quality is raised to the extent that the probability of a correct response based on the alphabetum is increased from .758 to .889, the advantage of lexical over alphabetic evidence decreases from 13.9% to −.5%.

When stimulus information is low (when only a few features are detected in each letter location), lexical knowledge should be more beneficial. However, when the subject has only partial information about each letter, Johnston's (1978) procedure for computing lexical constraint (based on complete knowledge of the three context letters and no information about the target) may no longer correlate with the lexical constraint provided by a partial set of features at each letter location. Our analysis completely supports this hypothesis: Johnston's high-constraint words yield a PPC of .830 compared to .852 for the low-constraint set. Furthermore, the average number of word units exceeding criterion is exactly the same (3.4) for both sets of words. It is clear that there is absolutely no relation between the number of letters that will form a word in the critical position of a test word (Johnston's definition of lexical constraint) and the number of words that are visually similar to that word (the candidate words in the activation–verification model).

In contrast, when lexical constraint is defined as the amount of lexical benefit, the effects of lexical constraint are apparent in the data. For each of the 288 stimuli of each type, we subtracted $P(C/A)$ from $P(C/L)$ and then partitioned the stimuli into thirds on the basis of these differences. For both sets of word data and the pseudoword data, obtained performance on the most highly constrained third is about 5% greater than that on the bottom third. There were no differences for the anagrams, but this is to be expected because our anagrams rarely activate the lexicon. Although the effect of lexical constraint (defined as lexical benefit) is small, it appears in all three data sets where it was predicted to occur. Furthermore, this measure provides a pure index of the predictive power of the lexical branch of our model. This is true because the psychophysical distinctiveness of the target letter is removed by subtracting $P(C/A)$. Differences in lexical constraint are due only to the mixture of candidate words that support the correct, incorrect, or neither alternative.

Another way of appreciating the role of lexical constraint in our data is to compare the high-constraint (top third) and low-constraint (bottom third) words to the high- and low-constraint anagrams. The magnitude of the WSE is about 10% for the high-constraint set compared to only 5% for the low-constraint set. One might speculate that a comparable effect of lexical constraint could be found in Johnston's (1978) data if they were analyzed on the basis of our new measure of lexical constraint.

Orthography

Massaro and his associates (Massaro, 1973, 1979; Massaro, Taylor, Venezky, Jastrzembski, & Lucas, 1980; Massaro, Venezky, & Taylor, 1979), have convincingly advocated a model in which letter recognition is guided by inferences drawn from knowledge of orthographic structure. Our model has no provision for the dynamic use of orthographic rules, nor does it assume a syllabary of commonly occurring letter clusters that could be activated by, or in parallel with, the alphabetum. Although it is clear that the model does not need any orthographic mechanism in order to predict the advantage of the regular pseudowords over the irregular anagrams, the present experiments offer a large set of stimuli and data to assess the possible contribution of orthography within the word, pseudoword, and anagram classes.

In accordance with the procedure advocated by Massaro, the sum of the logarithms of the bigram frequencies (SLBF) was computed for each stimulus. The correlations between SLBF and the two sets of word data were .11 and .04. Apparently, there is no relation between this measure of orthographic structure and performance on individual items. This is also true for the correlation between SLBF and the pseudoword data ($r = .09$). In contrast, the correlation between SLBF and the anagram data is much higher ($r = .30$). This pattern of correlation is similar to a previous analysis of orthographic structure (Paap & Newsome, 1980b) and further supports our conclusion that orthographic structure will predict performance only when very low levels are compared to somewhat higher levels of structure.

Although current data do not permit one to rule out the use of orthographic rules in letter and word recognition, our model shows that both the lexical (advantage of words over well-formed pseudowords) and orthographic (advantage of pseudowords over irregular strings) component of the WSE can be predicted on the basis of lexical constraint alone. Furthermore, lexical access may also account for the apparent effect of orthography on anagram performance. In the activation–verification model, the contribution of lexical activity is determined by the probability of a decision based on the lexicon, $P(L)$, and the probability of a correct response based on lexical activity, $P(C/L)$. The correlation between orthography (SLBF) for each anagram and its corresponding $P(L)$ is .49. Furthermore, the correlation between SLBF and $P(C/L)$ is also .49. In terms of our model, there is no direct effect of orthographic structure on letter recognition. Rather, it is simply the case that extremely irregular letter strings rarely excite the lexicon and, therefore, cannot benefit from lexical access. On the other hand, less irregular anagrams will occasionally activate a word unit, and that unit is likely to support the correct alternative.

Recently, Massaro (Note 7) conducted simulations of his fuzzy logical model that are similar to the activation–verification model in that top-down evidence (e.g., log bigram frequencies) is combined with an index of visual evidence based on letter-confusion probabilities. For six-letter anagrams visual evidence alone is a poor predictor; the correlation between predicted and observed results for 160 anagrams is only .08. Adding the log-bigram frequency component to the model raises the correlation to .59. Orthography does seem to have a considerable impact and suggests the possibility that perception of longer strings may be influenced by orthographic regularity to a much greater extent than is perception of shorter strings. On the other hand, it is entirely possible that the activation–verification model may also be able to account for the orthographic effects in Massaro's six-letter anagrams on the basis of lexical access and without recourse to any orthographic mechanism.

The outcome of Massaro's simulation for the 40 six-letter words is less informative. The correlation between obtained data and that predicted from the visual component alone was .48 compared to only .43 for the model that combines both the visual and orthographic components. This suggests that the impact of orthography on the perception of six-letter words may be quite weak, but it may be important to note that performance levels were not at all comparable for the words (90% correct) and anagrams (75% correct).

Comparisons of the Interactive Activation and Activation–Verification Models

McClelland and Rumelhart (1981; Rumelhart & McClelland, 1982) have proposed an interactive activation model that extends to the same wide scope of letter and word recognition paradigms that have been the target of our activation–verification model. Both models share many basic assumptions: (a) that stimulus input activates spatially specific letter units, (b) that activated letter units modulate the activity of word units, and (c) that letter and word recognition are frequently affected by important top-down processes. These generally stated assumptions permit both models to predict and explain the effects of lexicality, orthography, word frequency, and priming. However, the specific operations used to instantiate these general assumptions in McClelland and Rumelhart's computer simulation and in our computational algorithms offer a large number of provocative differences with respect to the specific mechanisms responsible for the various contextual phenomena. Furthermore, the two models are not always equally adept in accounting for the various context effects.

The Word and Pseudoword Advantage

The WSE is often characterized as consisting of two effects. The lexical effect refers to the benefits that accrue from accessing the lexicon and is estimated from the obtained advantage of words over well-formed pseudowords. The orthographic effect refers to the benefits derived from the reader's knowledge of orthographic redundancy and can be estimated from the obtained advantage of pseudowords over irregular nonwords. Both the activation–verification and interactive activation models assume that lexical activation accounts for both lexical and orthographic effects.

In the interactive activation model, lexical access facilitates letter recognition through excitatory feedback from activated word units to their constitutent letter units. Word stimuli are very likely to activate word units that reinforce the letters presented, thereby increasing the perceptibility of the letters. In contrast, irregular nonwords will rarely activate a word unit, and accordingly, the persistence of activity in the correct letters units will not be extended by feedback. Because pseudowords share many letters in common with words, they too activate word units that produce excitatory feedback and strengthen the letter units that give rise to them.

Given the detailed encoding assumptions of the interactive activation model and the particular set of parameter values needed to predict the basic pseudoword advantage, McClelland and Rumelhart conclude that the amount of feedback, and hence the amount of facilitation, depends primarily on the activation of word units that share three letters with the stimulus. They call the set of words that share three letters with the stimulus its neighborhood. The amount of facilitation for any particular target letter will be primarly determined by the number of word units in the neighborhood that support the target ("friends") and the number that support some other letter ("enemies").

This generalization provides a good basis for comparing the two models, because the amount of facilitation produced by lexical access in our model will be primarily determined by the number of friends and enemies in the candidate set generated by our encoding algorithm. The set of words in the neighborhood of a particular stimulus is likely to be quite different from the set of candidate words. One major reason for this (as pointed out earlier in the discussion of the geometric mean as a measure of word-unit activation) is that word units that share three letters with the stimulus will fail to exceed the word-unit criterion if the mismatching letter is not very confusable with the letter actually presented. For example, for the input string SINK with *S* as the test letter, our encoding algorithm generates only three friends (SING, SINE, and SINK) and four enemies (LINK, WINK, FINK, and RINK). In addition to all of these words, the neighborhood includes five new friends (SICK, SANK, SINS, SILK, and SUNK) and two new enemies (*PINK* and MINK). Thus, the ratio of friends to enemies is 3:4 for our model compared to 8:6 for their model.

Using the candidate set generated by our model and the neighborhood defined by a search of our lexicon (the 1,600 four-letter words in the Kucera and Francis, 1967, norms), we computed the proportion of friends for each stimulus according to each of the two models. In order to compare the predictive power of the two models, we then correlated the proportion of friends against the two sets of word data, the anagram data, and the pseudoword data. For all four cases the proportion of friends in the candidate set yielded higher correlations than the proportion of friends in the neighborhood. The average correlation for our model was .24 compared to .14 for the interactive activation model. In summary, our model seems to have a slight edge in its ability to account for consistent interitem variation that accrues from lexical access.

We were also curious as to the implications that McClelland and Rumelhart's encoding assumptions would have for the average performance on our words, pseudowords, and anagrams. To this end the alphabetic branch of our model was modified so that (a) the activity of each word was boosted by .07 for each matching letter and reduced by .04 for each mismatching letter and (b) the word-unit criterion would be exceeded by all those lexical entries that shared at least three letters in common with the stimulus. The first modification is based on the values of letter-to-word excitation and inhibition used by McClelland and Rumelhart and amounts to assigning a strength of .28 to the word unit corresponding to a word stimulus, and a strength of .17 to all the word units that share three letters with a stimulus. The probability of a decision based on the lexicon, $P(L)$, and the probability of a correct response based on lexical access, $P(C/L)$, were then computed as usual.

The decision rule was also the same, but deserves a brief comment. To extend McClelland and Rumelhart's analysis of the neighborhood to predictions of proportion correct in a two-alternative forced-choice task, it is necessary to separate nonaligned neighbors from true enemies. That is, word units in the neighborhood that support the incorrect alternative (true enemies) will have a much more disruptive effect on performance than words that support neither alternative (nonaligned neighbors). This is essentially what is done in Equation 2 for our model when we assume that friends contribute to a correct response with a probability of 1, nonaligned neighbors with a probability of .5, and true enemies with a probability of 0.

When a neighborhood based on the characteristics of the interactive activation model is substituted for the candidate set generated by our encoding algorithm, and all other operations are identical, the average predicted performance is .80 for words, .84 for pseudowords, and .74 for anagrams. This will not do at all, because the advantage of words over anagrams is too small and, more importantly, words are predicted to be inferior to pseudowords! McClelland and Rumelhart have already discussed why pseudowords tend to have a high proportion of friends. We add to their analysis a similar account of why words tends to have a lower proportion of friends.

Experimenters select stimulus words in

pairs that differ by only a single letter. This ensures that the two alternatives in the target location will both form words in the remaining context. For example, two of Johnston's (1978) high-constraint words were SINK and WINK, with the first position being probed with the alternatives S and W. One consequence of this is that every word stimulus will have at least one friend (itself) and one true enemy (its mate). Experimenters create pseudowords by substituting one of the context letters from the original word pair. For example, we created the pseudowords SONK and WONK by replacing the Is from SINK and WINK with Os. The consequence of this is that every pseudoword has at least one friend (SINK for SONK and WINK for WONK) but no built-in enemy (WONK is not an enemy of SONK because it is not a word). This systematic bias introduced in the selection of the materials results in the words' neighborhood averaging only 70% friends compared to 79% for the pseudowords. Thus, models based directly on the composition of the neighborhood will predict an advantage of pseudowords over words.

In fairness to the interactive activation model, it should be clearly pointed out that when its encoding assumptions are placed in the context of its own complete model, rather than our complete model, the simulation shows the correct ordering for the words, pseudowords, and single letters used by McClelland and Johnston (1977). We suspect that their full simulation would also produce the correct ordering of our words, pseudowords, and anagrams. The reason for this is that the complete interactive activation model assumes large (parameter value = .21) amounts of inhibition between competing word units. Thus, when a word is presented, the initial strength of the corresponding word unit (about .28) will quickly dominate the initial activity (about .17) of any potential enemy. Thus, the effects of lexical access for word stimuli are almost entirely determined by feedback from the corresponding word unit and no others. This is an interesting contrast between the two models. We assume that both the word advantage and the pseudoword advantage are mediated by decisions based on the activity of a small set of candidate words. McClelland and Rumelhart assume that the word advantage is mediated by feedback from a single word unit (the lexical entry corresponding to the word presented) but that the pseudoword advantage is mediated by feedback from large neighborhoods.

This inherent difference between words and pseudowords in the interactive activation model produces some undesirable fallout. Specifically, if high levels of interword inhibition permit the stimulus word to dominate any potential competition, then the stimulus-driven differences between various words will be eliminated. In short, high levels of interword inhibition mean that the functional amount of activation produced by the presentation of all words will be about the same. Thus, the significant correlations between obtained performance and that predicted from our model would stand unchallenged by the interactive activation model. It is true that the interactive activation model does predict some variation between words that is not stimulus driven, namely, that the resting levels of word units increase with word frequency, but we will show in a subsequent section that this assumption is not a good one.

Throughout the preceding section we have compared the predictive power of our model's candidate sets to that of McClelland and Rumelhart's neighborhood. Our encoding algorithm, which is highly sensitive to visual-confusability effects, seems to enjoy a consistent advantage in the tests we have conducted. However, this should not be viewed as a permanent disadvantage for the interactive activation model because the neighborhoods we tested conform to those obtained when their parameter, p, for visual-feature extraction is set to 1.0. If a value lower than 1.0 is used, their model will generate neighborhoods sensitive to visual confusability in a way similar to that of our candidate words. However, one of the difficulties in using the interactive activation model as a heuristic device is its inherent complexity. Accordingly, it is difficult to anticipate the results of simulations that have not been conducted. It should not be presumed in advance that the interactive activation model would accurately predict the relative differences between words, pseudowords, and anagrams when only partial information is gained from each letter location. Furthermore, when the contribution of visual confusability is introduced through the partial sampling of subjectively defined features it is not as likely to be as predictive as when confusability is based on an empirically derived confusion matrix.

The Pseudoword Expectancy Effect

One potential problem for any model that eschews any direct contribution of orthographic knowledge is that the pseudoword advantage seems to be more susceptible to expectancy effects than the word advantage. Carr, Davidson, and Hawkins (1978) have shown that if subjects do not expect to see any pseudowords, then performance on an unexpected pseudoword will be no better than that obtained with irregular nonwords. In contrast, they showed that the advantage of words over irregular nonwords was the same regardless of whether the subject expected all words or all nonwords.

McClelland and Rumelhart can account for this pattern of expectancy effects by assuming that subjects have strategic control over the degree of inhibition between the alphabetum and lexicon. They assume that if subjects expect only words or only irregular nonwords, they will adopt a large value of letter-to-word inhibition. More specifically, the inhibition parameter in their simulation is set so that the excitation produced by three matching letters will be precisely countered by the inhibition from the remaining mismatch. Accordingly, the only word unit that will produce appreciable feedback to the letter units is the word presented. This means that the word advantage will be about the same as always but that the pseudoword advantage will be eliminated.

Our activation–verification model can also predict the pseudoword expectancy results by assuming that subjects have control over one parameter, namely, the word-unit criterion. All of the predictions reported earlier used a word-unit criterion of .24. The average numbers of candidate words produced by the three classes of stimuli were 3.4 for words, 2.1 for pseudowords, and .2 for anagrams. By adopting this fairly lax criterion, the subject can take advantage of beneficial lexical evidence for both words and, more importantly, pseudowords. However, because the word unit corresponding to a word stimulus would exceed a much stiffer criterion, subjects have no motivation to maintain a low criterion and, therefore, to consider larger sets of word units unless they expect to see some pseudowords.

The expectancy effect was modeled by raising the word-unit criterion from .24 to .29. This resulted in a reduction of the number of candidate words to 1.4 for word stimuli, .40 for pseudowords, and .04 for anagrams. The effect of this on the predicted proportion correct is negligible for words (.841 versus .856) and anagrams (.755 versus .747) but results in a sizable decrease in pseudoword performance (.813 to .760). In summary, raising the word-unit criterion can result in the elimination of the pseudoword advantage while having very little effect on the word advantage. Although a higher criterion does lead to an increase in $P(C/L)$ for word stimuli, this tends to be countered by a decrease in the total amount of lexical activity and, hence, a decrease in $P(L)$.

Both models can predict the pseudoword expectancy effect reported by Carr et al. (1978). Although introspection is at best a weak test of two opposing theories, we yield to the temptation to point out that it seems to us more natural that a subject-controlled strategy might involve the adjustment of a criterion for considering lexical evidence rather than the adjustment of the amount of inhibition between letter and word detectors.

Word-Frequency Effects for Masked Stimuli

Under normal conditions of stimulus presentation, familiar words can be processed more effectively than less familiar ones. For example, high-frequency words are consistently classified faster than low-frequency words in lexical-decision tasks (Landauer & Freedman, 1968; Rubenstein, Garfield, & Millikan, 1970; Scarborough, Cortese, & Scarborough, 1977). Our complete model captures this familiarity effect by assuming that the order of verification is determined, in part, by word frequency. However, it was assumed that the brief stimulus durations used in the present experiments, together with the masking fields, would prevent verification from taking place.

Two studies have systematically manipulated word frequency under conditions of backward masking. In his first experiment Manelis (1977) selected 32-word sets from the Kucera and Francis (1967) norms with high (94–895), medium (23–74), and low (2–10) frequency counts. Although proportion of correct recognitions increased with frequency from .775 to .794 to .800, the differences were not significant. In the second experiment pairs of high- and low-frequency words shared the same critical letter and as many context letters as possible. Again, there were no differences between common (.762) and rare (.757) words. In a set of three experiments described by Paap and Newsome (1980b), 80 words were selected from the Thorndike-Lorge (1944) count so that there were equal numbers of words with frequencies of 1, 2, 5, 14, and 23 per million. Words in the five frequency classes were matched in terms of the identity and position of the target letter. The proportions of correct responses, in increasing order of frequency, were .67, .62, .65, .66, and .65.

The results described above support our assumption that verification does not occur when stimulus words are followed by a mask. We have also tested for word-frequency effects in the data we obtained with Johnston's (1978) words. The Kucera and Francis frequency counts were determined for each of the 288 words and correlated against both sets of word data. These correlations are shown in parentheses in Table 5. There are no significant correlations between word frequency and proportion correct, and in fact, the trend is toward poorer performance with higher word frequency. However, when a logarithmic transformation is applied to the frequency counts, positive correlations appear in each of the data sets.

Because many of Johnston's (1978) words are quite uncommon and may not be entered in the subjective lexicon of our typical subject, it is possible that this small word-frequency effect reflects nothing more than the probability of the word appearing in the lexicon. This interpretation was investigated by sequentially removing the words with the lowest frequency from the original set of 288 words. As shown in Table 5, the correlation between the logarithm of word frequency and performance systematically decreases as rare words are removed from the sample. When only words with frequencies greater than three are considered, there is no effect of relative frequency.

In order to further support our claim that many of Johnston's (1978) words are unfamiliar to our population of undergraduate subjects, we had 147 students classify each of the words as either (a) a word that I know the meaning of, (b) a word that I don't know the meaning of, or (c) a nonword. Thirteen words were classified as nonwords by a majority of the subjects (LAVE, TING, BOON, CRAG, WHET, JILL, BOLL, WILE, HONE, HEWN, FIFE, BANS, VATS). Furthermore, for many words the responses were distributed quite evenly across the three categories (e.g., FIFE, BANS, VATS, TEEM, HEMP, PENT, WANE, NAVE, SLAT). When we removed the 35 words that are most often classified as nonwords (and the meaning of which is known by only a minority of the subjects), there were no significant correlations between the data for the individual words and the logarithm of their frequency. This purging of our lexicon also led to a slight improvement in the correlation between predicted and obtained performance for the 288 words, $r = .32$.

These tests lead us to conclude that masking almost always prevents verification and that there is no need to build word-frequency effects into our encoding algorithm. In order to make sure that word frequency could not enhance the ability of our encoding algorithm to predict variation between words, we tried several different ways of having the logarithm of word frequency modulate the activity of the word units. Our basic strategy, like that of McClelland and Rumelhart, was to decrease the stimulus-driven activity of word units in inverse relation to their frequency. Because the correlation between our obtained word data and log word frequency was .16, we searched for a frequency effect that would produce a comparable correlation between our predicted data and log word frequency. The desired impact of word frequency was achieved when the amount of stimulus-driven activity was reduced by about 5% for each half-log unit drop in word frequency. This means that the most common words in our lexicon would receive no reduction in activity, and those with a frequency of only one would be reduced by 40%.

Because the word-frequency effect leads to an overall reduction in lexical activity, it was necessary to lower the word-unit criterion substantially (.14) in order to maintain candidate sets of about 3.3 words. Under these conditions the predicted performance for all words was exactly the same (PPC = .84) as that predicted from the original model that has no provision for word-frequency effects. The question of interest can now be answered: Does word frequency enhance the model's ability to account for variation between words? No, the correlations between predicted data and two sets of obtained data show that introducing word-frequency effects produces no change for one data set and a decline of .06 for the other.

In summary, we can find no evidence in our data or elsewhere that two-alternative forced-choice performance on masked word displays shows a word-frequency effect. This is consistent with the activation–verification model, because we assume that word frequency does not affect activation of the word units, but will affect the order of verification when the stimulus-presentation conditions permit verification to occur. The magnitude of the word-frequency effects generated by the interactive activation model is not known. Although their model specifically assumes that the resting activity of word units is determined by familiarity, other factors, such as the decision rules adopted for the forced-choice task, may severely attenuate the initial frequency differences between word units and, thereby, permit the prediction of no word-frequency effect. A fair conclusion with respect to word frequency is that the activation–verification model can correctly predict the magnitude of familiarity effects in both tachistoscopic and reaction time studies and that the interactive activation model may be able to do so.

Reaction Time Studies

As we mentioned in the introduction, the concepts embodied in our activation–verification model were originally developed in the context of reaction time studies using lexical-decision and naming tasks. With this history it is to be expected that the model can handle a variety of reaction time data. There are too many findings to cover in detail here, but it may be useful to review some of this earlier work to provide some idea about the performance of the model. Because the interactive activation model has not been specifically applied to lexical-decision data, we cannot draw specific comparisons. However, the interactive activation model has been used to explain the effects of semantic context and word frequency in other reaction time tasks (e.g., naming tasks), and we will comment on the applicability of analogous explanations of findings from the lexical-decision task.

The interactive activation model and our activation–verification model differ about the nature of effects of prior semantic context and word frequency when stimuli are pre-

Table 5
Correlations Between Obtained Proportion Correct and Log Word Frequency

Data set	Word frequencies included: All	All > 1	All > 2	All > 3
Word set				
1	.16 (−.04)	.14 (−.06)	.09 (+.04)	.04 (−.08)
2	.14 (−.09)	.11 (−.11)	.07 (+.01)	.04 (−.14)
Number of words	288	249	228	220
$r = .05$	.12	.13	.13	.13

Note. Correlations between proportion correct and the absolute word-frequency counts are shown in parentheses. "All > 1" means all stimulus words with a frequency greater than 1.

sented for normal durations and without masking. In the interactive activation model, these two factors both have the effect of increasing activation levels in relevant word units. The base activation level of the word units increases as a function of word frequency. Also, word units that are related to the context have increased activity levels relative to word units for unrelated words. Perhaps word units that are inconsistent with the context would have depressed activity levels as well.

In contrast, our activation–verification model places the effects of word frequency subsequent to the activation of word units. Word frequency determines the order in which lexical units are verified in the verification process. The activation–verification model also assumes that context increases the activity level of lexical units that are related to the context, but this activity increase may be high enough to cause the word units to exceed the criterion for inclusion in the candidate set. The verification process is then responsible for the analysis of stimulus information. Thus, verification can prevent a premature response. There appears to be no comparable mechanism in the interactive activation model.

In lexical-decision tasks, there is evidence that context and frequency have different effects on the time required to classify a letter string as a word. Becker and Killion (1977) found that context interacts with the quality of the visual stimulus whereas frequency and visual quality show additive effects. These findings imply that frequency and context exert their influence on performance in different ways, contrary to expectations, derived from the interactive activation model. McDonald (1980) developed a computer simulation of the verification model (which was the precursor to our activation–verification model). McDonald's simulation produced both the additivity of frequency and visual quality and the interaction of context and visual quality. Further, as we discussed earlier, there are apparently no word-frequency effects in the word-superiority paradigm. This result follows naturally from our model because frequency does not affect the activation process, which is the basis of the decision in the word-superiority paradigm.

The activation–verification model is also consistent with findings on effects of context on the classification of nonwords in the lexical-decision task. Several models (including the interactive activation model) handle context effects by inducing a bias in favor of related words. This approach leads to the expectation that nonwords that are very similar to particular words should be erroneously classified as words more often in a related context than in an unrelated context. For example, the nonword NERSE should be misclassified more often following a word related to NURSE (e.g., DOCTOR) than following an unrelated word (e.g., LAMP). In contrast, our model assumes that lexical decisions are made on the basis of verification rather than activation and that the quality of the verification process is not affected by context. Context affects the availability of lexical units for verification, but not the quality of the verification process itself. Thus, context should have no effect on the liklihood of classifying a nonword as a word.

The evidence on the classification of nonwords supports the predictions of the activation–verification model. Schvaneveldt and McDonald (1981) found no effect of context on classifying nonwords when stimuli remained available until the response occurred. Context did facilitate response time to words in their experiments. Other studies have produced similar results (Antos, 1979; Lapinski, 1979; McDonald, 1977, 1980; Lapinski & Tweedy, Note 8). O'Connor and Forster (1981) concluded that a bias explanation was ruled out by their findings even though one of their experiments showed bias effects. In that experiment, however, error rates were over 35% on the critical items, which is unusually high. In the context of the activation–verification model, such error rates suggest that subjects are responding without verification on a substantial proportion of the trials. If verification is optional, speed–accuracy trade-offs may be partly due to the probability of verification in a particular task. Schvaneveldt and McDonald (1981) also showed bias effects of context with a brief stimulus display followed by a masking stimulus. As we argued earlier, we assume that these stimulus conditions prevent verification.

Overall, the activation-verification model appears to handle a considerable amount of data from reaction time experiments (see Becker, 1980, and McDonald, 1980, for further examples). We believe that one important characteristic of the model lies in the independent top-down analysis of the stimulus (verification) that is sensitive to deviations from the stored representation of a word. These deviations might be further divided into permissible (identity preserving) and illegal (identity transforming) distortions of the stored representation. Verification, then, amounts to determining whether the stimulus impinging on the senses could be reasonably interpreted as a particular word after context or the senses had suggested that the stimulus might be that word.

We have presented our solution to what we perceive as an important theoretical problem in pattern-recognition theory in general and word recognition in particular. That problem is to specify the nature and interaction of bottom-up and top-down information-processng activities in recognition. There seems to be wide acceptance of the necessity for both of these types of processes. There is less agreement about just what they are and how they interact. Our solution to this theoretical problem provides a top-down process that involves comparing stimulus information to prototypes stored in memory. As such, the top-down process may enhance perception of discrepancies rather than induce a perceptual or decision bias in favor of expected stimuli. We believe that the evidence supports our view, but we are eager to pursue the matter further with additional research. We hope that our theoretical analysis and the contrasts of two theoretical approaches will help to focus further experimentation.

Reference Notes

1. Paap, K. R., & Newsome, S. L. The role of word-shape and lexical constraint in the word superiority effect. In C. Cofer (Chair), *Some new perspectives on word recognition.* Symposium presented at the meeting of the Southwestern Psychological Association, Houston, April 1981.
2. Paap, K. R., & Newsome, S. L. *Lexical constraint: Redefined and resurrected.* Paper presented at the meeting of the Psychonomic Society, Philadelphia, November 1981.
3. Paap, K. R., Newsome, S. L., & McDonald, J. E. *Further tests of the contribution of perceptual confusions to the WSE.* Paper presented at the meeting of the Psychonomic Society, St. Louis, November 1980.
4. Schvaneveldt, R. S., & McDonald, J. E. The verification model of word recognition. In C. Cofer (Chair), *Some new perspectives on word recognition.* Symposium presented at the meeting of the Southwestern Psychological Association, Houston, April 1981.
5. Becker, C. A., Schvaneveldt, R. W., & Gomez, L. *Semantic, graphemic, and phonetic factors in word recognition.* Paper presented at the meeting of the Psychonomic Society, St. Louis, November 1973.
6. Paap, K. R., Newsome, S. L., McDonald, J. E., & Schvaneveldt, R. W. *The activation verification model: The effects of cuing, masking, and visual angle.* Manuscript in preparation, 1982.
7. Massaro, D. W. *Simulating letter and word recognition: A fuzzy logical model of integrating visual information and orthographic structure in reading.* Paper presented at the European Conference on Artifical Intelligence, Orsay, France, July 1982.
8. Lapinsky, R. H., & Tweedy, J. R. *Associate-like nonwords in a lexical-decision task: Paradoxical semantic context effects.* Paper presented at the Mathematical Psychology meetings, New York University, August 1976.

References

Antos, S. J. Processing facilitation in a lexical decision task. *Journal of Experimental Psychology: Human Perception and Performance,* 1979, *5,* 527–545.

Becker, C. A. Allocation of attention during visual word recognition. *Journal of Experimental Psychology: Human Perception and Performance,* 1976, *2,* 556–566.

Becker, C. A. Semantic context effects in visual word recognition: An analysis of semantic strategies. *Memory and Cognition,* 1980, *8,* 493–512.

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

Carr, T. H., Davidson, B. J., & Hawkins, H. L. Perceptual flexibility in word recognition. Strategies affect orthographic computation but not lexical access. *Journal of Experimental Psychology: Human Perception and Performance,* 1978, *4,* 674–690.

Johnston, J. C. A test of the sophisticated guessing theory of word perception. *Cognitive Psychology,* 1978, *10,* 123–153.

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

Landauer, T., & Freedman, J. Information retrieval from long-term memory: Category size and recognition time. *Journal of Verbal Learning and Verbal Behavior,* 1968, *7,* 291–295.

Lapinsky, R. H. *Sensitivity and bias in the lexical decision task.* Unpublished doctoral dissertation, State University of New York at Stony Brook, 1979.

Luce, R. D. A threshold theory for simple detection experiments. *Psychological Review,* 1963, *70,* 61–79.

Manelis, J. Frequency and meaningfulness in tachistoscopic word perception. *American Journal of Psychology*, 1977, *99*, 269–280.

Massaro, D. W. Perception of letters, words, and nonwords. *Journal of Experimental Psychology*, 1973, *100*, 349–353.

Massaro, D. W. Letter information and orthographic context in word perception. *Journal of Experimental Psychology: Human Perception and Performance*, 1979, *5*, 595–609.

Massaro, D. W., Venezky, R. L., & Taylor, G. A. Orthographic regularity, positional frequency, and visual processing of letter strings. *Journal of Experimental Psychology: General*, 1979, *108*, 107–124.

Massaro, D. W., Taylor, G. A., Venezky, R. L., Jastrzembski, J. E., & Lucas, P. A. *Letter and word perception: Orthographic structure and visual processing in reading.* Amsterdam: North Holland, 1980.

McClelland, J. L., & Johnston, J. C. The role of familiar units in perception of words and nonwords. *Perception & Psychophysics*, 1977, *22*, 249–261.

McClelland, J. L., & Rumelhart, D. E. An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review*, 1981, *88*, 375–407.

McDonald, J. E. *Strategy in a lexical decision task.* Unpublished master's thesis, New Mexico State University, 1977.

McDonald, J. E. *An information processing analysis of word recognition.* Unpublished doctoral dissertation, New Mexico State University, 1980.

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

O'Connor, R. E., & Forster, K. I. Criterion bias and search sequence bias in word recognition. *Memory & Cognition*, 1981, *9*, 78–92.

Paap, K. R., & Newsome, S. L. Do small visual angles produce a word superiority effect or differential lateral masking? *Memory & Cognition*, 1980, *8*, 1–14. (a)

Paap, K. R., & Newsome, S. L. A perceptual-confusion account of the WSE in the target search paradigm. *Perception & Psychophysics*, 1980, *27*, 444–456. (b)

Rubenstein, H., Garfield, L., & Millikan, J. Homographic entries in the internal lexicon. *Journal of Verbal Learning and Verbal Behavior*, 1970, *9*, 487–494.

Rumelhart, D. E., & McClelland, J. L. An interactive activation model of context effects in letter perception: Part 2. The contextual enhancement effect and some tests and extensions of the model. *Psychological Review*, 1982, *89*, 60–94.

Scarborough, D. L., Cortese, C., & Scarborough, H. S. Frequency and repetition effects in lexical memory. *Journal of Experimental Psychology: Human Perception and Performance*, 1977, *3*, 1–17.

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

Schvaneveldt, R. W., Meyer, D. E., & Becker, C. A. Lexical ambiguity, semantic context, and visual word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, 1976, *2*, 243–256.

Thorndike, E. L., & Lorge, I. *The teacher's word book of 30,000 words.* New York: Teacher's College Press, 1944.

Received October 21, 1981
Revision received January 28, 1982 ■