Journal of Experimental Psychology:
Learning, Memory, and Cognition
1996, Vol. 22, No. 1, 63-85

0278-7393/96/$3.00

Semantic Ambiguity Effects in Word Identification

Ron Borowsky and Michael E. J. Masson
University of Victoria

The influence of semantic ambiguity on word identification processes was explored in a series of word naming and lexical-decision experiments. There was no reliable ambiguity effect in 2 naming experiments, although an ambiguity advantage in lexical decision was obtained when orthographically legal nonwords were used. No ambiguity effect was found in lexical decision when orthographically illegal nonwords were used, implying a semantic locus for the ambiguity advantage. These results were simulated by using a distributed memory model that also produces the ambiguity disadvantage in gaze duration that has been obtained with a reading comprehension task. Ambiguity effects in the model arise from the model's attempt to activate multiple meanings of an ambiguous word in response to presentation of that word's orthographic pattern. Reasons for discrepancies in empirical results and implications for distributed memory models are considered.

Any comprehensive theory of mental representation and process must accommodate the complex means by which concepts are communicated through language. Through the course of history, humans have developed tools of communication that facilitate the relaying of ideas and concepts, such as a writing system or *orthography*. This mapping of concepts to orthography is not entirely one to one, however, resulting in some words that correspond to multiple concepts, which are known as *semantically ambiguous words*. When reading text, the context provided by preceding words and sentences provides a means of disambiguating such words. As a result, we may not even notice the ambiguity in words that we are reading in context. If, on the other hand, semantically ambiguous words are presented in isolation, their alternative meanings are readily accessible, and thus their ambiguous nature is noticed. In the research reported in this article, we compare performance on semantically ambiguous words with that of semantically unambiguous words in isolated word identification tasks and describe simulations of the empirical effects within the framework of a distributed memory architecture (Masson, 1995).

The effect of semantic ambiguity on isolated word identification has usually been determined by comparing performance on unambiguous words (which are associated with only one meaning) to performance on ambiguous words (which are associated with more than one meaning). There are many reports of an advantage in response latency for ambiguous as compared with unambiguous words, both in the word naming task, in which participants are required to pronounce words (e.g., Balota, Ferraro, & Conner, 1991; Fera, Joordens, Balota, Ferraro, & Besner, 1992; Hino & Lupker, 1993), and in the lexical-decision task, in which participants are asked to decide whether letter strings spell words (e.g., Jastrzembski, 1981; Jastrzembski & Stanners, 1975; Kellas, Ferraro, & Simpson, 1988; Millis & Button, 1989; Pugh, Rexer, & Katz, 1994; Rubenstein, Garfield, & Millikan, 1970; Rubenstein, Lewis, & Rubenstein, 1971).

Models of Semantic Ambiguity Effects

An exploration of how the language processing system handles semantically ambiguous words may be particularly informative with respect to theories of lexical representation and processing. Two classes of visual word identification models have been used to examine the processing advantage for semantically ambiguous words: localist and distributed representation models. Localist representation models assume that lexical information is represented in specific units that correspond to individual words. One type of localist representation model is based on the principle of *serial search* (e.g., Forster & Bednall, 1976; Rubenstein et al., 1970). In these models, orthographic input resulting from the presentation of a word is compared with a set of lexical entries one at a time, with the search terminating when the correct entry is located. This process occurs in two stages. First, the letters that make up the word are identified, with the representation of this information serving as an access code for selecting a subset of lexical entries. Second, a serial search (ordered by word frequency) of this subset is carried out until a match is made and verified against the orthographic input. The ambiguity advantage in serial search models stems from ambiguous words having separate lexical entries and the greater probability that one of these multiple correct entries will be accessed

Ron Borowsky and Michael E. J. Masson, Department of Psychology, University of Victoria, British Columbia, Canada.

This research was supported by the Natural Sciences and Engineering Research Council of Canada. We thank Steve Lindsay for use of his equipment. We are grateful to Derek Besner and Paul Fera for suggesting that neighborhood density be examined in these experiments and to Rick Bourassa for suggesting the delayed naming experiment.

Correspondence concerning this article should be addressed to Ron Borowsky, who is now at the Department of Psychology, 9 Campus Drive, University of Saskatchewan, Saskatoon, Saskatchewan, Canada S7N 5A5, or to Michael E. J. Masson, Department of Psychology, University of Victoria, P.O. Box 3050, Victoria, British Columbia, Canada V8W 3P5. Electronic mail may be sent via Internet to Ron Borowsky at ron.borowsky@usask.ca or to Michael E. J. Masson at mmasson@sol.uvic.ca.

during verification as compared with the case of an unambiguous word for which there would be only one correct entry.

Activation models constitute another type of localist representation model (e.g., the Jastrzembski, 1981, extension of Morton's, 1969, logogen model and the Kellas et al., 1988, extension of McClelland and Rumelhart's, 1981, interactive activation model). These models are founded on the assumption that the lexical system consists of one detector, or *logogen,* for each word or concept known to the reader. These detectors serve as evidence collectors when orthographic input is presented, and a word is recognized when the evidence accrued on its behalf reaches a threshold. Multiple detectors can collect evidence at the same time, and it is assumed in these models that there are separate detectors corresponding to the separate meanings of ambiguous words. The ambiguity advantage in activation models thus stems from the greater probability of an ambiguous word activating one of its multiple detectors compared with an unambiguous word activating its only detector. This advantage has been described further within an interactive activation framework as being due to greater inhibition of other competing detectors by the multiple detectors corresponding to an ambiguous word, which in turn serves to feed a greater amount of converging evidence (and thus facilitation) back to a letter detector level (Kellas et al., 1988), thereby enhancing word identification.

The second class of model that has been applied to the problem of understanding semantic ambiguity effects consists of distributed representation models (known also as *parallel distributed processing* or *distributed memory models*). Distributed representation models represent lexical knowledge in weights associated with links that connect a set of processing units to one another and instantiate a known word by evoking its unique pattern of activation across the processing units. In contrast to localist representation models, there is no single processing unit that corresponds to a known word. The representation and processing of semantically ambiguous words present a challenge for distributed representation models because one orthographic pattern must be mapped onto two different patterns of activation among the collection of units that represent meaning. At an intuitive level, this mapping problem implies that an ambiguity disadvantage, not an advantage, should be observed in word identification tasks.

Of the distributed representation models that have been used to simulate word identification processes (e.g., Hinton & Shallice, 1991; Joordens & Besner, 1994; Kawamoto, 1993; Kawamoto, Farrar, & Kello, 1994; Masson, 1991, 1995; Plaut & Shallice, 1993), there are only two accounts of the semantic ambiguity advantage (Joordens & Besner, 1994, and Kawamoto et al., 1994). We now turn to a brief description of these two accounts.

Using the distributed memory model developed by Masson (1991), Joordens and Besner (1994) attempted to simulate the ambiguity advantage in lexical decision. This model consists of two processing modules: one representing the orthography of words, and the other representing the meanings of words. Joordens and Besner found that after learning ambiguous words (orthographic patterns that are mapped onto two different meaning patterns on different learning trials), the model often failed to settle into one of the appropriate meaning patterns of an ambiguous word. Instead, the model settled into a *blend,* representing a mixture of the two learned meaning patterns. If these blend states are considered to represent errors, as Joordens and Besner have argued, then the main problem is a simulated error rate (74% errors) that is far from being comparable with human performance (usually less than 15% errors). In terms of accuracy then, this model clearly does not account for the empirical data. However, when they examined the simulated lexical-decision response latency for words (nonwords were not presented to the network in their simulations), as measured by the number of processing cycles for the meaning module to settle on a correct pattern, an ambiguity advantage was found.

The basis for the ambiguity advantage found by Joordens and Besner (1994) was a proximity effect associated with the meaning units. At the beginning of a word identification trial, meaning units were placed into a random pattern. By chance, such a starting pattern is likely to be closer to one of the two meanings of an ambiguous word than to the only meaning of an unambiguous word. When the starting pattern is closer to the target pattern into which the meaning units must settle to complete a lexical decision, fewer processing cycles are required to move the units into the target pattern.

The model developed by Kawamoto et al. (1994) provides a different account of the ambiguity advantage in the lexical-decision task. Their model consists of an orthographic and a meaning module and uses an error-correction learning algorithm (unlike the Hebbian learning rule used by Joordens and Besner [1994] and in the simulations reported later in this article). We do not present the details of their model here, but some aspects of it are germane to our discussion on semantic ambiguity. In the case of an ambiguous word, the model is presented with an inconsistent mapping between an orthographic pattern and a meaning pattern. The connection weights between orthographic and meaning units therefore do not receive consistent modifications across learning trials that involve the different meanings of the ambiguous word. To compensate for this inconsistency when learning ambiguous words, the error-correction learning algorithm makes the connection weights between orthographic units—which take on the same pattern of activation on all learning trials involving an ambiguous word—particularly strong. For an unambiguous word, the consistent mapping between orthography and meaning leads the learning algorithm to generate more moderate connection weights both within and between modules.

To simulate the lexical-decision task, Kawamoto et al. (1994) assumed that the orthographic units must settle into a stable pattern of activation. The orthographic units settled more quickly for ambiguous words because the connection weights between orthographic units were more strongly influenced by the orthographic patterns of these words than for unambiguous words.

Robustness of Empirical Semantic Ambiguity Effects

How confident can researchers be in the empirical effects of ambiguity that have been the target of these recent simulation efforts? Several researchers have raised this concern (e.g., Joordens & Besner, 1994; Masson & Borowsky, 1995; Rueckl, 1995), and rightly so. The between-stimuli nature of the ambiguity advantage requires that some caution be exercised when considering its validity. Indeed, there are a substantial number of reports in which no such advantage was found in the lexical-decision task (Clark, 1973; Forster & Bednall, 1976; Gernsbacher, 1984). As is usually the case with between-stimuli effects, accounts for discrepancies in the empirical results regarding semantic ambiguity have typically emphasized the artifactual influence of some confounding variable (such as differences in subjective familiarity, for example, Gernsbacher, 1984), discrepancies between laboratories regarding how the variable of interest (i.e., ambiguity) is measured (e.g., Kellas et al., 1988; Millis & Button, 1989; Rubenstein et al., 1971), or the conservative nature of the statistical tests used (Clark, 1973; Forster & Bednall, 1976). Thus, it is difficult to assess the validity of a between-stimuli effect like that of semantic ambiguity.

Although the generality of an effect such as an ambiguity advantage will always be compromised by its between-stimulus nature, steps can be taken to maximize the validity of the effect. Some of the potentially confounding (or *extraneous*) variables can be controlled by matching the two sets of words on such variables or by partialling out their influence by using multiple regression techniques. Neither of these approaches alone is likely to be adequate. Although items might be matched on a number of extraneous variables, a sizeable portion of variability in task performance may be determined by these factors. If that variability is not accounted for (e.g., as a factor in the experiment), it remains as part of the error term used in assessing the effect of the factor of interest. In the multiple regression approach, the degree of multicolinearity between extraneous variables and the variable of interest will restrict the amount of unique variance in the dependent variable that can be accounted for by the variable of interest. Using these techniques together, however, generates a number of benefits. Matching stimuli as closely as possible on extraneous variables legitimizes the use of the more powerful repeated measures analysis of variance as opposed to a between-groups analysis. In the multiple regression analysis, matching serves to orthogonalize extraneous variables with respect to the variable of interest when they are regressed on the dependent variable. These variables will then not compete for the same variance in the dependent variable. Instead, extraneous variables will account for variability that is unrelated to the variable of interest, making for a more powerful test of the variable of interest's unique predictive strength.

In Experiment 1 we used materials taken from another study in which an ambiguity advantage was reported (Fera et al., 1992), but in subsequent experiments we adopted the approach of matching items to permit the application of a repeated measures analysis by items and to produce a more powerful regression analysis. We report the results of these experiments in which reliable evidence for an ambiguity advantage was found in the lexical-decision task but not in the naming task. A modified version of the distributed memory model described by Masson (1995) is introduced and simulations that replicate the observed pattern of ambiguity effects are presented.

Experiment 1

In our first attempt at examining potential semantic ambiguity effects in naming, we used the same materials as Fera et al. (1992). In selecting their materials, Fera et al. defined ambiguous words in the same way as Kellas et al. (1988; Ferraro & Kellas, 1990). That definition was based on a task in which participants decided whether a given word has one or more than one meaning. A word was considered ambiguous if a sufficient percentage of the participants claimed it had more than one meaning.

The procedure we used for the word naming task was identical to that of Fera et al. (1992, Experiment 1), with the following exceptions. First, rather than having participants code their own errors, an experimenter was present during testing to code the responses. Second, rather than having each trial begin at a fixed time after the preceding trial, participants controlled the onset of each trial.

Method

Participants. Thirty University of Victoria students participated in the experiment for extra credit in an introductory psychology course. All participants had normal or corrected-to-normal vision and considered English to be their first language.

Materials and design. The critical stimuli consisted of the 60 unambiguous and 60 ambiguous words used by Fera et al. (1992). Fera et al. defined ambiguity for these stimuli by using the same procedure as Kellas et al. (1988). The two sets of critical words were equal with respect to mean word frequency (43.4 per million; Kučera & Francis, 1967). In addition to the critical words, we used 16 practice words. Half of the practice words were unambiguous and half were ambiguous words according to the Ferraro and Kellas (1990) ratings, with the exception of 2 words that did not appear in the Ferraro and Kellas ratings.

Apparatus. Micro Experimental Laboratories (MEL) software and an IBM-compatible computer controlled the stimulus displays and the timing of events as well as recorded the data. The stimuli appeared on two monochrome monitors; participants viewed the stimuli on one monitor while the experimenter observed an identical display on the other monitor. A microphone was connected to the MEL button box–voice key apparatus to detect the onset of vocalization. Participants controlled the rate of stimulus presentation by pressing the spacebar on the computer keyboard, and the experimenter coded the accuracy of the participant's response by using the MEL button box. Response latency was measured from the onset of the target on the screen to the onset of the participant's vocalization.

Procedure. Participants were individually tested in a quiet room. The procedure lasted approximately 15 min. Participants sat in front of a monitor and were instructed, both in writing and verbally, that they would see one word on each trial and that they should pronounce each word as quickly and as accurately as possible. The experimenter coded each response *as correctly pronounced, incorrectly pronounced,* or *spoiled* (i.e., stutter, failed to trigger voice key, or other noise-triggered voice key). Participants initiated stimulus presentation by pressing the spacebar on the computer keyboard.

The sequence of events was as follows: (a) a fixation row of three asterisks appeared in the center of the screen, (b) the participant pressed the spacebar to initiate the trial, (c) there was a 275-ms interstimulus interval, (d) a word appeared in lowercase letters and remained visible until the participant responded, and (e) the screen was blank for a variable amount of time during which the experimenter coded the correctness of the response.

Results

Following Fera et al. (1992), we based response latency analyses on correct responses after trimming latencies that fell outside the range of 150 to 1,200 ms. This criterion excluded only 0.03% of correct trials, allowing us to retain almost all of the data. The mean and median response latencies for remaining correct trials were computed for each participant (averaging across items) and for each item (averaging across subjects). Unlike previous studies, in which the effects of semantic ambiguity on mean response latency have been assessed, analyses of both mean and median response latencies were carried out. We emphasize here analyses of median response latencies, however, because the median is less affected than the mean by outliers. An analysis of variance (ANOVA) of mean response latency is reported only if it differed from that of median response latency. A trial was classified as an error if the participant mispronounced the word, accidentally triggered the voice key, or if the voice key failed to detect the participant's response. The error rate for each subject and item was computed as the percentage of trials on which an error was made. The means of subject and item latency means and medians and the mean for subject and item error percentages are shown in Table 1.

The latency and error data were submitted to two types of ANOVA, with ambiguity as the variable. One analysis was a repeated measures ANOVA, with subjects as the random variable, and the other was an independent-groups ANOVA, with items as the random variable. Unless otherwise stated, the Type I error rate was .05. There was no effect of ambiguity on median response latency by subjects or by items (Fs < 1). The analysis of mean response latency yielded a reliable ambiguity effect by subjects, $F(1, 29) = 4.61$, $MSE = 84.63$, but not by items ($F < 1$). The power to detect an ambiguity effect of the size obtained by Fera et al. (1992; effect size was 13 ms by subjects and 19.5 ms by items) on median response latency was estimated as .98 by subjects and .91 by items. The analyses of error percentages also failed to reveal an ambiguity effect by subjects, $F(1, 29) = 1.95$, $MSE = 5.34$, or by items, $F(1, 118) = 1.07$, $MSE = 19.56$.

Table 1
Mean and Median Latencies (in Milliseconds) and Percentage of Error Rates as a Function of Ambiguity in Experiments 1 and 2

	Mean latency		Median latency		Error rate (%)	
Experiment	U	A	U	A	U	A
Experiment 1						
Subjects						
M	508	503	495	494	4.8	4.0
SD	59	59	56	61	4.3	2.9
Items						
M	507	503	491	489	4.8	3.9
SD	31	23	34	29	4.7	4.1
Experiment 2						
Subjects						
M	495	494	485	485	5.1	4.1
SD	44	45	42	45	2.6	2.9
Items						
M	495	494	484	485	5.1	4.1
SD	21	19	23	21	5.1	3.9

Note. U = unambiguous; A = ambiguous.

To determine whether Experiment 1 was sufficiently sensitive to detect the effects of word attributes other than ambiguity, we computed correlations across items between median response latency and orthographic neighborhood density (N, which represents the number of words [including the target word] orthographically similar to the target word as determined by changing one letter at a time; Coltheart, Davelaar, Jonasson, & Besner, 1977), length, number of higher frequency neighbors (NHF, which has been shown to facilitate naming of low-frequency words, see Grainger, 1990), phonological error score (PE, from the Seidenberg & McClelland, 1989, parallel distributed processing model of word recognition, which can be used as a rough measure of spelling-to-sound regularity such that the lower the phonological error score, the higher the spelling-to-sound regularity and word frequency), orthographic error score (OE, also from the Seidenberg and McClelland model, 1989, which can be used as a rough measure of orthographic familiarity and word frequency such that the lower the error score, the higher the orthographic familiarity), and word frequency (from Kučera and Francis, 1967, which was transformed to $\log_{10}$ word frequency to create a linear relationship between response latency and word frequency; see Balota & Chumbley, 1984; Borowsky & Besner, 1993). The correlations between these variables, including ambiguity (coded as one for unambiguous words and two for ambiguous words), are presented in Table 2. The correlations between these variables and the dependent measures (median response latency and error percentage) are shown in Table 3. Response latency was significantly correlated with N, length, PE, OE, and word frequency.

As a further test for an ambiguity effect, we conducted multiple regression analyses. The rationale for these analyses was that a small ambiguity effect might be masked by the relatively large amount of variability in response latency or error rates. Thus, we reasoned that if variability that was due to extraneous factors could be removed from response latencies and error rates, then a small ambiguity effect might be detected.

As a first step in setting up the regression analyses, we examined the stimulus variables that we had measured to find ones that might serve as unique predictors of latency. Note that, in Table 2, each variable is significantly correlated with at least two other variables. In selecting predictors for inclusion in the regression analysis, we followed two guidelines. First, ambiguity, N, NHF, PE, and OE were to be included. Ambiguity was included as the variable of interest, and the remaining variables were included because they were confounded (i.e., correlated) with ambiguity. Second, other variables significantly correlated with the dependent measure (median response latency or error rate), but not with each other, were included. Whenever a potential predictor variable was found to be correlated with another predictor variable, only the variable that was most strongly related to the dependent measure was included in the analysis. We adopted this approach to avoid multicolinearity among predictor variables.

The correlations between the set of potential predictor variables and median response latency and percentage of error

Table 2
Correlations Between the Measures of Word Characteristics in Experiment 1 and Experiments 2–4

Characteristic	Ambiguity	N	Familiarity	Length	NHF	PE	OE	Frequency
Ambiguity	—	.34*		−.12	.21*	−.23*	−.40*	.18
N	.14	—		−.42*	.69*	−.21*	−.38*	.02
Familiarity	−.10	.10	—					
Length	.04	−.63*	−.08	—	−.30*	.01	.21*	.10
NHF	.02	.75*	−.15	−.53*	—	−.11	−.16	−.10
PE	−.01	−.22*	−.29*	−.01	−.07	—	−.21*	−.50*
OE	.03	−.30*	−.43*	.04	−.13	.49*	—	−.47*
Frequency	.00	−.12	.58*	.24*	−.41*	−.46*	−.53*	—

Note. Correlations above the diagonal are for the words used in Experiment 1. (Blank cells indicate that familiarity ratings were not collected in Experiment 1.) Correlations below the diagonal are for the words used in Experiment 2–4. In Experiment 1, for correlations involving PE or OE, n = 127, otherwise n = 128. In Experiments 2–4, for correlations involving PE or OE, n = 119, otherwise n = 120. N = orthographic neighborhood density; NHF = number of higher frequency neighbors; PE = phonological error score; OE = orthographic error score.
*$p < .05$.

are shown in Table 3. The correlations in Table 3 indicate that N, length, PE, OE, and frequency were significantly related to latency. Length and frequency, then, are the only potential predictor variables other than those that were deemed necessary to include in the regression equation (i.e., ambiguity, N, NHF, PE, and OE). Because length and frequency were not correlated with one another, they were both chosen for inclusion in the regression analysis.

A simultaneous multiple regression was conducted on latency, with ambiguity, N, NHF, PE, OE, length, and frequency as predictors. A summary of the regression analysis can be seen in Table 4. The unique relation between latency and ambiguity (i.e., after removing variance that was attributable to N, NHF, PE, OE, length, and frequency) was not significant. Both length and frequency accounted for reliable, unique amounts of latency variability. N did not account for a reliable amount of unique variability because N and length were strongly related in this sample of words and length had a stronger relationship with latency.

Table 3
Correlations Involving Median Latencies and Percentage of Error Rates in Experiments 1 and 2 With Word Characteristics

	Experiment 1[a]		Experiment 2[b]	
Characteristic	Latency	Error rate (%)	Latency	Error rate (%)
Ambiguity	−.03	−.10	.03	−.04
N	−.30*	−.07	−.24*	.14
Familiarity	—	—	−.39*	−.12
Length	.46*	.03	.45*	−.05
NHF	−.15	−.04	−.18*	.20*
PE	.21*	.01	.25*	.10
OE	.30*	.04	.14	−.11
Frequency	−.21*	−.08	−.10	−.17

Note. Familiarity ratings were not collected in Experiment 1. N = orthographic neighborhood density; NHF = number of higher frequency neighbors; PE = phonological error score; OE = orthographic error score.
[a]For correlations involving PE or OE, n = 119, otherwise n = 120. [b]For correlations involving PE or OE, n = 123, otherwise n = 124.
*$p < .05$.

There were no variables significantly related to error percentage, so only the variables that were confounded with ambiguity were entered as predictors in a simultaneous multiple regression, with error percentage as the criterion variable. The results of this analysis are shown in Table 4. None of the predictors accounted for a reliable amount of unique variability in error percentage.

Table 4
Summary of Regression Analyses in Experiments 1 and 2

Criterion and predictor variables	Coefficient	t	df
	Experiment 1[a]		
Latency			
Ambiguity	8.85	1.60	112
N	−1.01	−1.52	112
NHF	.75	.58	112
PE	.65	.44	112
OE	.71	.64	112
Length	16.99	4.66***	112
Frequency	−11.16	−2.27*	112
Error rate (%)			
Ambiguity	−.68	−.73	114
N	−.03	−.28	114
NHF	.01	.04	114
PE	−.04	−.16	114
OE	.01	.05	114
	Experiment 2[b]		
Latency			
Ambiguity	−1.56	−.47	119
N	.44	1.14	119
Familiarity	−7.99	−4.85***	119
Length	13.65	5.00***	119
Error rate (%)			
Ambiguity	−.38	−.51	120
N	.00	.02	120
NHF	.26	1.42	120

Note. N = orthographic neighborhood density; NHF = number of higher frequency neighbors; PE = phonological error score; OE = orthographic error score.
[a]n = 120. [b]n = 124.
*$p < .05$. ***$p < .001$.

Discussion

The ambiguity effect on response latency (by subjects) was significant when mean latencies were computed for each participant, but not when median response latencies were used, nor when item analyses were done. This pattern of results suggests that the ambiguity effect by subjects may have been due to outliers in the response latency distributions. Given that the outlier trimming procedure we used removed only 0.03% of the data, this account is plausible. It is not clear, however, why Fera et al. (1992), using the same materials and outlier trimming procedure with latency data of similar magnitude, obtained a reliable ambiguity effect when they analyzed subject and item means. Our regression analyses also showed no significant effect of ambiguity on naming performance but did reveal that our experiment was sensitive enough to detect significant unique effects of word length and frequency on median response latency.

One possibility is that the two classes of words differed on some other factor to which participants in the Fera et al. (1992) experiment were more sensitive than were our participants. For example, although mean word frequency was equal for the ambiguous and unambiguous words, median frequency was 20 for the ambiguous words and 10.5 for the unambiguous words. Thus, it appears that these two groups of words were not closely matched on word frequency. Second, there was also a slight difference with respect to word length. Ambiguous words were shorter than unambiguous words (*M*s = 4.3 letters vs. 4.6 letters). For example, there were eight 3-letter words in the ambiguous word set, and only one 3-letter word in the unambiguous word set. Given the demonstration of a positive correlation between word length and median response latency in Experiment 1, differences in length may be important. Third, N has been suggested by D. Besner and P. Fera (personal communication, February 14, 1994) to be a confounding factor with these stimuli. Indeed, N was found to be significantly greater among ambiguous words than among unambiguous words (*M*s = 10.9 vs. 7.4). Given that Andrews (1989, 1992) found shorter response latencies in both lexical decision and naming for words with high N (although Andrews's effect appears to be restricted to low-frequency words) and given our own finding that N was significantly correlated with naming latency, N must be considered as a potential confound when examining the effect of semantic ambiguity. Fourth, NHF, PE and OE were found to be significantly related to ambiguity, and, thus, they must also be considered as potentially confounding variables.

Experiment 2

Given the concerns about possible differences between the ambiguous and unambiguous words used by Fera et al. (1992) and in Experiment 1, we developed a set of unambiguous words that were better matched to the ambiguous stimuli and used these items in a second experiment. We attempted to match stimuli with respect to word length, word frequency, N, NHF, PE, and OE, as well as initial phoneme. Thus, for each semantically ambiguous word used by Fera et al., an unambiguous word was chosen to create a pair of words that was as closely matched as possible on these seven variables. This type of pairwise matching also legitimizes a more powerful repeated measures analysis to be conducted when treating items as a random factor.

Subjective familiarity ratings were also collected. Gernsbacher (1984) has demonstrated that a confound between familiarity and ambiguity can produce an apparent ambiguity advantage in lexical decision, underscoring the importance of using stimuli that are equated on this variable. Furthermore, Gernsbacher has convincingly argued that familiarity ratings are a more sensitive measure of the actual frequency of encounters with words and pointed out that familiarity ratings are more contemporary than printed frequency counts.

Method

Participants. The participants were 42 students drawn from the same population as in Experiment 1. All participants had normal or corrected-to-normal vision and considered English to be their first language.

Materials and procedure. The same apparatus, materials, design, and procedure used in Experiment 1 were used in Experiment 2, with the following exceptions. Four additional semantically ambiguous words from Experiment 2 of Fera et al. (1992) were included, resulting in a total of 64 ambiguous words. Sixty-four unambiguous words were selected with the constraint that they have only one entry in the *Random House Dictionary* (1978) and that they not appear as an ambiguous word (i.e., did not have a mean ambiguity rating greater than 1.5) in the Ferarro and Kellas (1990) corpus. These 64 unambiguous words were matched as closely as possible to the ambiguous words on initial phoneme, word length, and word frequency. In selecting these unambiguous items, we also considered N, NHF, PE and OE (see the Appendix).[1]

After participating in the naming task, 19 of the 42 participants made familiarity judgments on a 7 point scale with responses ranging from 1 (*very unfamiliar*) to 7 (*very familiar*). Participants were asked to base their decisions on how often they could remember having seen or heard the word. Participants were encouraged to use all of the numbers between one and seven in making their judgments. The familiarity rating for each item and the mean over all items within each group are shown in the Appendix. The mean familiarity rating was similar for ambiguous and unambiguous words, $F(1, 63) = 1.87$, $MSE = 0.70$.

Repeated measures ANOVAs applied to each of the measures in the Appendix indicated that the sets of words were comparable on these variables (*F* ratios did not approach significance), except for N, which was significantly higher for ambiguous words, $F(1, 63) = 4.98$, $MSE = 15.70$. As noted earlier, N has a facilitative effect on naming latencies (Andrews, 1989, 1992), so any difference that was due to this variable should favor ambiguous word naming latencies.

Results

As in Experiment 1, latencies were included in the analysis only if the naming response was correct and the latency fell within the range of 150 to 1,200 ms. This constraint excluded 0.08% of correct trials. The means of subject and item mean and median response latencies and the means of subject and

[1] Summed bigram frequency was also examined but did not differ between the two groups of words, nor did it correlate with any of the dependent variables in any of the experiments.

item error percentages are presented in Table 1. Both measures of average latency revealed only very small differences between ambiguous and unambiguous words.

The same analyses used in Experiment 1 were conducted on the data from Experiment 2, except that repeated measures analyses were used when treating items as a random variable because the items in Experiment 2 were matched pairs. There was no ambiguity effect on median response latency by subjects or by items ($Fs < 1$). Power to detect an ambiguity effect on median response latency of the size obtained by Fera et al. (1992) was estimated to be greater than .99, both by subjects and by items. The analysis of error percentages yielded a significant effect of ambiguity by subjects, $F(1, 41) = 4.40$, $MSE = 4.46$, but not by items ($F < 1$), with a higher error rate among unambiguous items.

Given the significant ambiguity effect on error percentage, it is possible that a speed–accuracy trade-off may have obscured an ambiguity effect on latency. On the other hand, the fact that the effect on errors did not hold in the items analysis led us to suspect that the effect in the subjects analysis was due to a few error-prone unambiguous words. To check this possibility we examined the item means for outliers. Two unambiguous items produced unusually high error percentages: *plug* and *steak*, with means of 19.0% and 23.8%, respectively. The experimenter noted that participants tended to pronounce *plug* as "plunge" (one of the practice items) and that *steak* tended to elicit the regularized pronunciation "steek." It is likely that participants regularized the pronunciation of *steak* because of some intrusions from the pronunciations of the seven other words in the stimulus set that contained *ea* (pronounced as /i/) in their word bodies *fear, lean, meat, stream, seat, steam,* and *tea* (see Burt and Humphreys, 1993, for a detailed discussion of such intrusion effects in naming). Because it was plausible that these outliers were responsible for the ambiguity effect on error percentage in the analysis by subjects, data for these two items and data for their matched ambiguous words were excluded, and the analyses of latencies and errors were recomputed. The latency analyses produced exactly the same results, including the power estimates. The mean error percentages (4.2% vs. 4.6% for ambiguous and unambiguous words, respectively) were not reliably different by subjects or by items ($Fs < 1$).[2] The data from these four items were also excluded from the analyses that follow.

In setting up the regression analyses, we examined the stimulus variables that we had measured to find ones that might serve as unique predictors of latency. The correlations between these variables, including ambiguity (coded as one for unambiguous words and two for ambiguous words), are presented in Table 2. Note that with the exception of ambiguity, each variable was significantly correlated with at least three other variables. By virtue of our matching procedure, these variables were orthogonal to ambiguity status in this set of items. In selecting predictors for inclusion in the regression analysis, we followed two guidelines. First, ambiguity and N were to be included. Ambiguity was included as the variable of interest, and N was included because there was a tendency for ambiguous words to have higher values of N and we were concerned about a potential confound. We note, however, that the correlation between ambiguity and N was not significant with these materials (see Table 2), although with a more sensitive repeated measures ANOVA, a reliable relationship was found. Second, other variables significantly correlated with the dependent measure (median response latency or error rate), but not with each other, were included. As described in Experiment 1, if a potential predictor variable was found to be correlated with another predictor variable, only the variable that was most strongly related to the dependent measure was included in the analysis to avoid multicolinearity among predictor variables.

The correlations between the set of potential predictor variables and median response latency and percentage of error are shown in Table 3. The correlations in Table 3 indicate that, ambiguity and N aside, four variables were significantly related to latency: familiarity, length, NHF, and PE. Familiarity and length were not reliably correlated with each other, but PE was related to familiarity, and NHF was related to length (see Table 2). Because familiarity had a stronger relationship with latency than did PE, and because length had a stronger relationship with latency than NHF, familiarity and length were chosen for inclusion in the regression analysis.

A simultaneous multiple regression was conducted on latency, then, with ambiguity, N, familiarity, and length as predictors. A summary of the regression analysis can be seen in Table 4. The unique relation between latency and ambiguity (i.e., after removing variance that was attributable to N, length, and familiarity) was not significant. Both length and familiarity accounted for reliable unique amounts of latency variability. N did not account for a reliable amount of unique variability because N and length were strongly related in this sample of words and length had a stronger relationship with latency.

The only factor related to error percentage was NHF so that variable was included with ambiguity and N in a simultaneous multiple regression, with error percentage as the dependent variable. The results of this analysis are shown in Table 4. None of the predictors accounted for a reliable amount of unique variability in error percentage.

Discussion

The results of Experiment 2 replicated those of Experiment 1. Using unambiguous words that were matched to the ambiguous words on initial phoneme, word frequency, and word length and that were similar with respect to spelling-to-sound regularity, orthographic familiarity, and number of higher frequency neighbors, we found no reliable ambiguity effects in naming latency. Moreover, correlational analyses

[2] We also noticed that six of the words used in Experiment 2 were not perfectly matched on word length (five ambiguous words were one letter longer than their partners, and the reverse was true for another pair; see the Appendix). Given that response latency and word length were related, we considered the possibility that absence of an ambiguity processing advantage in the present experiment may have been due to the unambiguous stimuli that were shorter than their ambiguous counterparts. Thus, the data from the six pairs of stimuli that were not perfectly matched on word length, and the two pairs that contained the error-prone words, were removed and the analyses were recomputed. The analyses produced the same results as when only the two error-prone items and their partners were removed.

indicated that response latency was sensitive to neighborhood density, familiarity ratings, word length, and phonological error score but not to ambiguity. Despite the correlation between latency and neighborhood density, and the larger average neighborhood density among ambiguous words in our sample, no ambiguity advantage in response latency was found. Furthermore, there was no significant unique effect of ambiguity on latency (partialling out variance attributable to neighborhood density, familiarity, and length). The absence of a significant unique relationship between ambiguity and response latency or error rate, coupled with finding no difference on naming latency or error rate when ambiguity was treated as a repeated measures factor, provides converging evidence for the conclusion that there is no advantage for ambiguous words in the naming task.

One reason for not being able to detect an ambiguity advantage in naming performance could be that an ambiguity disadvantage occurs during the production of a naming response to an ambiguous word. Although we had no a priori reason to hypothesize that such a disadvantage would occur during response production, the combination of an ambiguity advantage in the processes that lead up to response production, coupled with an ambiguity disadvantage in response production, would serve to nullify an ambiguity advantage in the naming task. Response production effects in naming can be isolated with a delayed naming task that requires the participant to delay the response until a cue appears (e.g., Balota & Chumbley, 1985; Monsell, Doyle, & Haggard, 1989). If a cue delay is used that is sufficiently long to allow all preproduction processes enough time to finish, then the resulting latency and error rate data will represent naming production performance independent of preproduction performance. If an ambiguity disadvantage during response production was the reason that a reliable ambiguity advantage was not found in Experiment 2, then a reliable ambiguity disadvantage in delayed naming performance should be seen. We carried out such an experiment only to find no sign of an ambiguity disadvantage. Thus, we can conclude that the lack of an ambiguity advantage in standard naming is not attributable to an opposing ambiguity processing disadvantage during naming production.

Experiment 3

Experiments 1 and 2 failed to replicate the ambiguity advantage reported by other researchers using the naming task (Balota et al., 1991; Fera et al., 1992; Hino & Lupker, 1993). Because the purported ambiguity advantage in naming did not appear to be robust, we also examined the effect in the lexical-decision task. Earlier demonstrations of an ambiguity advantage in the lexical-decision task (e.g., Kellas et al., 1988; Millis & Button, 1989; Pugh et al., 1994) might also be questioned on the grounds that some unidentified variable was confounded with ambiguity and was responsible for the effect. This possibility is quite real given that Gernsbacher (1984) showed that a number of earlier demonstrations of an ambiguity advantage in the lexical-decision task were compromised by a confound with the rated familiarity of the words.

In Experiment 3, we used the matched set of unambiguous and ambiguous words from Experiment 2 in a lexical-decision task. The nonwords in this experiment were pronounceable and orthographically legal. The data were analyzed by using both ANOVA and multiple regression, as in Experiments 1 and 2.

Method

Participants. The participants were 30 students drawn from the same population as in Experiments 1 and 2. Some students received $5 for their participation, and others received extra credit in an introductory psychology course. All participants had normal or corrected-to-normal vision and considered English to be their first language.

Materials and procedure. The same apparatus, materials, design, and procedure used in Experiment 2 were used in Experiment 3, with the following exceptions. Participants viewed stimuli on a color monitor and responded by pressing one of two keys (*z* and slash keys on the computer keyboard) to indicate their lexical-decision response. The key under the participant's dominant hand was used to signal *word,* and the key under the nondominant hand was used to signal *nonword.* Response latency was measured from the onset of the target on the screen to the participant's key press response. A set of 128 pronounceable nonwords (matched in length to the word stimuli) was added to the stimulus set, resulting in an equal number of word and nonword stimuli. Experimental trials were preceded by 16 practice trials (4 ambiguous words, 4 unambiguous words, and 8 nonwords). The experiment lasted approximately 25 min. Participants were instructed, both in writing and verbally, that they would see one letter string on each trial and that they should decide as quickly and as accurately as possible whether it spells a word that they know.

Results

Analyses of response latency were based on correct responses after removing response latencies that fell outside the range of 150–2,000 ms. This criterion excluded 0.23% of correct trials and allowed us to retain most of the data. One of the unambiguous words, *rev,* elicited an extremely high error rate (87% errors) and thus was removed from the analyses along with its ambiguous word partner, *row.* The means of individual subject and item mean and median response latencies, along with mean percentage of errors, are presented in Table 5.

The same analyses used in Experiment 2 were applied in Experiment 3. The ANOVA of median response latency yielded a significant ambiguity advantage by subjects, $F(1, 29) = 5.10$, $MSE = 248.27$, but not by items, $F(1, 62) = 2.19$, $MSE = 1{,}541.86$. The analysis of percentage error showed no significant ambiguity advantage in either the subjects analysis, $F(1, 29) = 2.95$, $MSE = 6.27$, or the items analysis, $F(1, 62) = 1.68$, $MSE = 23.12$.

The correlations between the measured stimulus variables and median response latency and error percentage were computed across the 126 critical items (excluding the *rev–row* pair). These correlations are presented in Table 6. Ambiguity was not significantly correlated with response latency or percentage of error. A multiple regression analysis, with response latency as the dependent variable, was conducted by following the same procedure as in Experiments 1 and 2. By those criteria, the predictor variables included in the regression analysis were ambiguity, N, and familiarity. Although frequency, OE, and PE were also correlated with response

Table 5
Mean and Median Lexical Decision Latencies (in Milliseconds) and Percentage of Error Rates as a Function of Ambiguity in Experiments 3 and 4

	Mean latency			Median latency			Error rate (%)		
Experiment	U	A	NW	U	A	NW	U	A	NW
Experiment 3									
Subjects									
M	647	637	729	610	601	687	4.3	3.2	6.4
SD	95	92	108	88	79	96	2.7	3.5	5.2
Items									
M	649	639	732	613	603	685	4.3	3.2	6.4
SD	49	51	71	50	44	64	6.6	3.5	7.9
Experiment 4									
Subjects									
M	567	569	570	540	540	546	2.6	2.3	2.3
SD	65	64	55	54	52	46	3.0	2.1	1.7
Items									
M	567	569	570	537	539	544	2.6	2.3	2.3
SD	34	34	34	28	28	31	3.3	3.1	3.7

Note. U = unambiguous; A = ambiguous; NW = nonword.

latency, all were significantly correlated with familiarity, which had the strongest correlation with latency.

A summary of the regression analysis can be seen in Table 7. The unique relation between latency and ambiguity (i.e., after removing variance that was attributable to N and familiarity) was significant. The negative coefficient indicates that, after removing any effects of familiarity and N, ambiguous words were, on average, responded to 17 ms faster than unambiguous words. The unique relation between latency and familiarity (after removing variance attributable to ambiguity and N) was also significant. There was no significant unique relation between N and latency.

A similar regression analysis was conducted with percentage of error as the dependent variable. Our procedure for selecting predictor variables resulted in the inclusion of ambiguity, N, familiarity, and NHF in the set of predictor variables. A summary of this analysis is shown in Table 7. As in the regression analysis of latency data, only ambiguity and familiarity accounted for reliable amounts of unique variance.

Table 6
Correlations Involving Median Latencies and Percentage of Error Rates in Experiments 3 and 4 With Word Characteristics

	Experiment 3		Experiment 4	
Characteristic	Latency	Error rate (%)	Latency	Error rate (%)
Ambiguity	−.11	−.11	.03	−.04
N	−.08	.03	−.01	.00
Familiarity	−.53*	−.47*	−.24*	−.10
Length	.05	−.11	.10	−.02
NHF	.16	.19*	.04	.00
PE	.21*	.26*	.18*	.00
OE	.35*	.24*	.13	−.11
Frequency	−.48*	−.40*	−.27*	−.05

Note. For correlations involving PE or OE, n = 125, otherwise n = 126. N = orthographic neighborhood density; NHF = number of higher frequency neighbors; PE = phonological error score; OE = orthographic error score.
*$p < .05$.

Table 7
Summary of Regression Analyses in Experiments 3 and 4

Criterion and Predictor variables	Coefficient	*t*	*df*
	Experiment 3		
Latency			
Ambiguity	−17.21	−2.41*	122
N	−.12	−.19	122
Familiarity	−26.87	−7.32***	122
Error rate (%)			
Ambiguity	−1.83	−2.13*	121
N	.02	.14	121
Familiarity	−2.61	−5.57***	121
NHF	.18	.77	121
	Experiment 4		
Latency			
Ambiguity	.00	.00	122
N	.03	.07	122
Familiarity	−6.98	−2.75**	122
Error rate (%)			
Ambiguity	−.27	−.47	123
N	.00	.08	123

Note. n = 126. N = orthographic neighborhood density; NHF = number of higher frequency neighbors.
*$p < .05$. **$p < .01$. ***$p < .001$.

Discussion

Although the ambiguity advantage in latency was significant by ANOVA only in the subjects analysis, a reliable items effect was obtained in the multiple regression analysis. In contrast to naming performance (Experiments 1 and 2), ambiguity accounted for significant unique variance in lexical-decision latency. This combination of findings suggests that semantic ambiguity produces an advantage in lexical decision but not in naming.

If we had not collected familiarity ratings, and had included frequency instead of familiarity as a predictor variable, would the ambiguity advantage still be reliable? Apparently not, the regression of ambiguity, N, and frequency on latency in Experiment 4 yielded only a significant unique effect of frequency (−43.67 ms per log unit of frequency), $t(122) = -6.30$, $p < .001$. The contrasting results obtained by using familiarity versus frequency attest to the fragility of the ambiguity advantage we have observed. If an insufficient proportion of variability in latency is accounted for by extraneous variables (e.g., familiarity), the proportion of variability accounted for by ambiguity may fail to reach significance.

Experiment 4

After finding an ambiguity advantage in lexical decision but not in naming, we sought evidence concerning the locus of the ambiguity advantage. Two views regarding the source of the advantage have been developed in the context of connectionist models of word identification. Joordens and Besner (1994) proposed that the ambiguity advantage is due to the distributed representation of the meaning of ambiguous words, thereby implying a *conceptual* locus; whereas Kawamoto et al. (1994) proposed that the advantage is due to stronger connection weights between orthographic units that are active in the representation of ambiguous words, thus implying an *orthographic* locus.

As an initial test of these ideas we conducted another lexical-decision experiment, this time using nonwords that could be distinguished faster and more easily from words. Orthographically legal nonwords (i.e., pronounceable nonword letter strings that are word like in their orthographic characteristics, such as those used in Experiment 3) are difficult to discriminate from words. When the task involves rejecting such nonwords, a more careful familiarity assessment or a more thorough search of memory for relevant conceptual information may be required than is the case when the nonwords are easily distinguishable at the orthographic level. The nonwords from Experiment 3 were modified by replacing vowels with consonants, thereby producing orthographically illegal nonwords (i.e., unpronounceable consonant strings). If the ambiguity advantage occurs at an orthographic level in word identification, then it ought to be present when the stimuli encourage lexical decisions to be primarily made on the basis of orthographic information. On the other hand, if the ambiguity effect has a semantic basis, then encouraging participants to make decisions at an orthographic level should reduce or eliminate the effect.

One effect of moving the basis of the lexical decision to the orthographic level was expected to be a reduction in response latency relative to Experiment 3. Another expected effect, given the reduced role of semantic processing, was that response latency for nonwords should more closely approximate latency for words (i.e., reducing the effect of lexicality). The expected similarity in latencies contrasts with the typical result, obtained in Experiment 3, in which response latency for nonwords is much longer than for words. Following this logic, we predicted an interaction between lexicality (word vs. nonword) and experiment (Experiment 3: legal nonwords vs. Experiment 4: illegal nonwords). If the locus of the semantic ambiguity advantage is, indeed, at the semantic level of processing, then we should also find an interaction between ambiguity and experiment in a regression analysis.

Method

Participants. The participants were 30 students drawn from the same population as in the earlier experiments. Some students received $5 for their participation, and others received extra credit in an introductory psychology course. All participants had normal or corrected-to-normal vision and considered English to be their first language.

Materials and procedure. The same apparatus, materials, design, and procedure used in Experiment 3 were used in Experiment 4, with the following exceptions. A set of 128 orthographically illegal nonwords was created from the orthographically legal nonwords used in Experiment 3 by replacing all vowels with consonants that resembled the vowels in shape.

Results

Analyses of response latencies were based on correct responses after excluding latencies that fell outside the range of 150–2,000 ms. This criterion excluded 0.12% of the correct trials. Data for the word *rev* and its ambiguous partner *row* were removed for the purpose of making the present analyses comparable to those of Experiment 3. The means of subject and item mean and median response latencies, along with mean percentage of errors, are shown in Table 5.

The same analyses used in Experiment 3 were used in Experiment 4. The analyses of median response latency yielded no effect of ambiguity by subjects or by items ($Fs < 1$). The analysis of percentage error also showed no sign of an ambiguity effect ($Fs < 1$). Power to detect an ambiguity effect on median response latency of the size obtained by subjects in Experiment 3 (i.e., 9 ms) was estimated to be .63.

A consistent (although nonsignificant) pattern of an ambiguity advantage on error percentages in Experiments 1, 2, and 4 caused us some concern in that an ambiguity advantage may have been emerging in error rates rather than in response latencies in these experiments. To test this possibility, we analyzed the error data from Experiments 1, 2, and 4 together. There was no significant ambiguity advantage on error percentage by subjects (unambiguous = 4.1% errors; ambiguous = 3.6% errors), $F(1, 101) = 2.18$, $MSE = 5.06$, or by items (unambiguous = 4.0% errors; ambiguous = 3.5% errors), $F(1, 368) = 1.30$, $MSE = 16.13$.

Regression analyses. The relationship between median response latency and error percentage and the set of stimulus characteristics explored in the earlier experiments was examined by computing correlations across the 126 critical items. These correlations are shown in Table 6. There was no significant correlation between ambiguity and either latency or error rate. As in the earlier experiments, however, multiple regression analyses were also conducted. Applying the same criteria for inclusion of predictor variables as in Experiments 2 and 3 led to a set of three predictors for response latency: ambiguity, N, and frequency. Because frequency and familiarity were so similar in the size of their correlations with latency, however, and because of our interest in comparing the results of the regression on Experiment 3 latency (which included familiarity as a predictor) to the regression on Experiment 4 latency, familiarity was included in the regression model instead of frequency. A summary of the regression analysis can be seen in Table 7. Only familiarity accounted for a significant proportion of unique variance in latency.[3] Power to detect an ambiguity effect on median response latency of the size obtained in the regression analysis of Experiment 3 (i.e., 17.21 ms) was estimated to be .97. The smallest effect size for which this experiment had at least .80 power to detect was 12.36 ms.

When error rate was used as the dependent variable, our criteria for entering predictor variables selected only ambigu-

---

[3] Similar results were obtained when frequency was used as a predictor instead of familiarity. Frequency was the only significant unique predictor of latency (−14.13 ms per log unit of frequency), $t(122) = -3.10, p < .01$.

ity and N. A summary of the results of this analysis is shown in Table 7. Neither variable was uniquely related to error rate.

Comparison of Experiments 3 and 4. To test our prediction that the use of orthographically illegal nonwords would serve to make lexical decisions more rapid (and also less influenced by semantic processing), we examined the effects of lexicality (words vs. nonwords) and the type of nonword used in the experiment (Experiment 3: legal nonwords vs. Experiment 4: illegal nonwords; see Table 5). In the ANOVA by items, lexicality was treated as a between-items effect, and experiment was treated as a within-item effect. In the ANOVA by subjects, lexicality was treated as a within-subject effect, and experiment was treated as a between-subjects effect. A significant main effect of lexicality on median latency was obtained by items, $F(1, 252) = 93.76$, $MSE = 2{,}295.65$, and by subjects, $F(1, 58) = 91.15$, $MSE = 822.29$. A reliable main effect of lexicality on percentage of error was also obtained by items, $F(1, 252) = 6.71$, $MSE = 28.49$, and by subjects, $F(1, 58) = 9.25$, $MSE = 8.97$. These lexicality effects both indicated a significant advantage for word performance over nonword performance when collapsing the data over experiments. A significant main effect of experiment on median latency was found in the analysis by items, $F(1, 252) = 817.97$, $MSE = 1{,}726.71$, and by subjects, $F(1, 58) = 33.11$, $MSE = 9{,}639.30$. The main effect of experiment was also found to be significant on error percentage by items, $F(1, 252) = 31.39$, $MSE = 28.77$, and by subjects, $F(1, 58) = 19.17$, $MSE = 11.13$. Taken together, the significant main effect of experiment on both latency and error percentage indicated an advantage for performance when illegal nonwords are used over performance when legal nonwords are used. Most important, however, a reliable interaction between lexicality and experiment was found on median latency by items, $F(1, 252) = 91.57$, $MSE = 1{,}726.71$, and by subjects, $F(1, 58) = 52.81$, $MSE = 822.29$. This interaction was also significant on percentage of error by items, $F(1, 252) = 7.97$, $MSE = 28.77$, and by subjects, $F(1, 58) = 6.04$, $MSE = 8.97$. This interaction of lexicality and experiment on both latency and error percentage indicated that the lexicality effect (i.e., the advantage for words over nonwords) diminished when illegal nonwords were used.

We also analyzed the interaction between ambiguity and type of nonword on the premise that if the ambiguity effect occurs primarily at the level of conceptual processing, then it ought to diminish in size when the task involves less conceptual processing. Although this hypothesis received some support from the significant ambiguity effect in Experiment 3 and the nonsignificant effect in Experiment 4, we sought more convincing evidence in the form of an interaction. In our analysis of this interaction, we included the median latency data from both Experiments 3 and 4 as a repeated measure and regressed the same predictor variables on latency as those used in the separate regressions on latency in Experiments 3 and 4: ambiguity, N, and familiarity. In this regression and in the one that follows, only the interaction between experiment (i.e., nonword type) and ambiguity was of interest, and only the test of this effect is reported. The median latency analysis revealed a significant interaction between experiment and ambiguity, $F(1, 122) = 4.34$, $MSE = 1{,}033.09$, indicating that the size of the ambiguity advantage was significantly reduced by using illegal nonwords. This interaction can be visualized by comparing the coefficient for ambiguity as a predictor of latency in Experiment 3 (17.21 ms in Table 7) with the same coefficient in Experiment 4 (0.00 ms in Table 7).

These analyses were also conducted on the percentage of error data. In examining the interaction between ambiguity and experiment, we included the same predictor variables as were used in the regression analysis of error percentage in Experiment 3, namely: ambiguity, N, familiarity, and NHF. The interaction between experiment and ambiguity was not reliable, $F(1, 121) = 1.99$, $MSE = 14.24$.

Discussion

By using consonant strings as nonwords in a lexical-decision task, thereby encouraging participants to make more rapid decisions on the basis of orthographic information, we eliminated the ambiguity advantage that had been obtained in Experiment 3. This interaction between ambiguity and experiment (i.e., nonword type), combined with the significant interaction between lexicality and experiment, suggested that the locus of the semantic ambiguity advantage is, indeed, at the semantic level and not at the orthographic level, contrary to the view proposed by Kawamoto et al. (1994). We can be reasonably certain that the task encouraged an orthographic basis for making lexical decisions because of the interaction between lexicality and experiment. In Experiment 4, response latencies were very similar for words and nonwords, in contrast to the shorter latency for words than for nonwords observed in Experiment 3. When nonword letter strings are less word like, as they were in Experiment 4, a lexical decision can be made faster, and thus there is less opportunity for semantic processing to contribute to the discrimination between words and nonwords. Under these conditions, the ambiguity advantage disappears.

It might be argued, however, that participants in Experiment 4 could have carried out the lexical-decision task merely by looking for the presence or absence of vowels. Such decisions would not be "lexical" at all because looking for the presence or absence of vowels can be accomplished without accessing any lexical representations or making any familiarity judgment. It is unlikely that this strategy was used extensively because word-level characteristics such as familiarity and frequency were found to be significantly correlated with response latency, and familiarity was found to be a unique predictor of response latency. Furthermore, in a regression on latency that included frequency instead of familiarity, frequency was a significant predictor of unique latency variability (see footnote 4). If participants had merely been searching for vowels, no effect of word familiarity or frequency should have been obtained. This does not, however, eliminate the possibility that a mixture of a vowel searching and lexical-access processes was carried out by our participants.

To preclude the use of a vowel-searching strategy, orthographically illegal nonwords that contain vowels, but are still unpronouncable, could be used; Shulman, Hornak, and Sanders (1978) used such nonwords only to find that the effect in which they were interested (i.e., graphemic similarity between pairs of words on lexical-decision performance) was similar to the effect obtained when consonant strings served as nonwords. However, the fact that Shulman et al. also found a semantic priming effect when orthographically illegal nonwords (containing vowels) were used in a lexical-decision task speaks to the automatic nature of access to semantic information during lexical decision. Although it remains to be seen whether this semantic priming effect would be reduced when consonant strings are used as nonwords, or enhanced when orthographically legal nonwords are used, it is clear that it is nearly impossible to make lexical decisions solely on the basis of orthographic level information (contrary to Kawamoto et al., 1994). Thus, although our use of consonant strings as nonwords in Experiment 4 can only discourage (but not preclude) the use of semantic information during lexical decision, it is likely that, overall, less semantic information contributed to lexical decision than in Experiment 3 in which legal nonwords were used. This differential contribution of semantic information suggests that the semantic ambiguity effect has a semantic locus because the effect was much smaller (i.e., nonsignificant) in Experiment 4. Evidence supporting this conclusion was also provided by Pugh, Rexer, and Katz (1994), who found that the ambiguity advantage in lexical decision disappears when nonwords with few or no orthographic neighbors are used. If nonwords have few or no orthographic neighbors, then it follows that they are much less word like in their orthographic characteristics than if the nonwords have many orthographic neighbors.

Simulation of Results With a Distributed Memory Model

We simulated the results of the naming and lexical-decision experiments by using a version of the distributed memory model described by Masson (1991, 1995). This model is a Hopfield network (Hopfield, 1982) consisting of three modules or sets of processing units: orthographic, phonological, and meaning. Each processing unit takes on one of two possible values or states (± 1). A word is instantiated in the network as a pattern of activation ($+1$ and -1 states) across the entire set of processing units. Knowledge about words is stored in the weights of the connections between processing units. The processing units are fully interconnected, both within and between modules. Connection weights are determined by a simple learning rule derived from Hebb (1949). When learning a pattern of activation, the connection weight between any pair of units is incremented if the two units are in the same state and decremented if the units are in different states. In particular, the change in the connection weight between a pair of units is defined as

$$\Delta w_{ij} = n_i n_j,$$

where w_{ij} represents the connection weight between units i and j, and n_i and n_j represent the states (± 1) of units i and j when the learned pattern is instantiated. This rule captures the correlations between units that happen to hold across the entire set of patterns that the network learns.

To simulate word identification tasks, the network is presented with an orthographic pattern corresponding to a word's printed form. The construction of the word's phonological pattern and instantiation of its meaning are simulated by updating the states of the phonological and meaning units. Units are randomly selected for updating until some criterion is reached. Updating consists of computing the activation coming into a unit and changing its state according to a simple threshold function. Activation coming into a unit is computed as

$$a_i = \sum_{i \neq j} w_{ij} n_j,$$

where a_i represents the amount of activation directed to unit i. The received activation is transformed into an activation value by a threshold function:

$$\text{if } a_i > 0, \text{ then } n_i = 1, \text{ else } n_i = -1.$$

This process of updating units amounts to moving the network through the space of possible patterns of activation into a basin of attraction that, under ideal circumstances, corresponds to a learned pattern of activation (i.e., a known word). Thus, given the orthographic pattern of a word, the network is capable of instantiating the phonological code for and the meaning of that word.

The simulations reported here were based on the same model architecture and procedures as those described by Masson (1995), with a few exceptions that are explained at appropriate points. A number of general changes should be noted at the outset. First, the architecture used for the present simulations consisted of an orthographic module with 70 units, a phonological module with 40 units, and a meaning module with 140 units. This arrangement contrasts with the architecture used by Masson (1995) in which there were more orthographic than meaning units. This modification was made because we discovered that when the network was trained on ambiguous words, the patterns of activation corresponding to the meanings of those words did not form stable basins of attraction unless a large number of units were assigned to the meaning module. A pattern of activation represents the bottom of a stable basin of attraction in a Hopfield (1982) net if, when the pattern is instantiated, no updating of units will lead to a change in state. Thus, once the system is in such a pattern of activation, it cannot escape unless external input or some other process intervenes.

Second, rather than starting a word identification trial by instantiating a word's complete orthographic pattern into the model's orthographic units, these units were randomly sampled for updating along with units from the other modules. When updating an orthographic unit, however, its state was completely determined by external input (i.e., the defined orthographic pattern) and not by activation received from other units in the network. To ensure that the orthographic pattern would be instantiated rather quickly, we set the probability of sampling a unit from the orthographic module to .6. As in the simulations reported by Masson (1995), sampling occurred more frequently from the phonological than from the meaning units (probabilities of .3 and .1, respectively), reflecting the role of a phonological recoding route in the retrieval of a word's meaning (Lukatela & Turvey, 1994; Perfetti & Bell, 1991; Perfetti, Bell, & Delaney, 1988; Van Orden, Johnston, & Hale, 1988; Van Orden, Pennington, & Stone, 1990).[4]

Third, for the sake of efficiency, two alterations were made with respect to sampling units for updating. Rather than sampling units with replacement, units within a module were randomly selected for updating without replacement. Once every unit in a module had been sampled, all units were once again eligible for sampling. Sampling of units within a module ceased after an entire pass through the units failed to produce a change of state in any of the units. An uneventful pass through the units of a module indicated that the module had reached a stable state.

In each of the simulations reported below, the network was trained on a small vocabulary consisting of two unambiguous words and two ambiguous words. Each ambiguous word had two unrelated meanings associated with it. The pattern of activation for each word was randomly constructed as a sequence of ± 1 values across the network's units. The network was trained by first setting all connection weights to zero, then presenting each unambiguous word twice and each ambiguous word twice (once with each of the different meanings) by using the Hebbian learning rule described above. Thus, the two types of words were equated with respect to frequency of exposure to the orthographic patterns of the words. After training, the network was tested 40 times on each word. Each test was conducted by setting all units to a random starting value of ± 1, then updating the units until a specified criterion was reached (see the *Word Naming, Lexical Decision,* and *Gaze Duration During Comprehension* sections for details regarding criteria). We used the number of cycles to reach criterion as a measure of response latency under the assumption that number of updating cycles is a monotonic function of time. The training–test sequence was run 80 times with independently generated words used on each run.

Word Naming

Performance on the word naming task in Experiment 2 was simulated by updating the network until the phonological units settled into a stable state. We assumed that once a stable state was reached, the participant would be able to produce a vocal response. The number of updating cycles required to reach stability was taken as a measure of response latency.

In preliminary simulations of the naming task we found that because of the large number of meaning units, the influence of the meaning module on the phonological units was too strong and the influence of the orthographic units was too weak. In particular, the model consistently produced a disadvantage for ambiguous words in the naming task, reflecting dynamics in the meaning module that are discussed below. Given that no effect of ambiguity was observed in the naming task we reduced the influence of meaning units on phonological units in an attempt to suppress this effect.[5] A reduced influence was achieved by weighting the activation coming from meaning units to phonological units by a factor of .3.

On each test trial, the network was allowed a maximum of 315 cycles to reach a stable state in the phonological units. Trials in which stability was not achieved by that time were considered errors. The limit of 315 cycles was selected so that the error rates would be similar to those observed in Experiment 2. The orthographic units typically reached full instantiation of a word's orthographic pattern by 150 cycles, so by that time only phonological and meaning units were sampled. The mean number of cycles required to reach a stable phonological pattern and the mean percentage of errors for unambiguous and ambiguous words are shown in Table 8. Separate ANOVAs were used to test the difference in means for latency and for error rates. Neither test produced a significant difference ($F < 2$ for latency; $F < 1$ for errors).

Some insight into the behavior of the network can be obtained by examining Figure 1, which shows the mean activation of a word's orthographic, phonological, and meaning patterns across cycles. Activation is defined as the proportion of units in a module that is currently in a state that matches the word's pattern of activation. Thus, at the beginning of a trial, half of the units (by chance) are in the appropriate state. As updating progresses, more of the units are moved into the appropriate state. Activation values are plotted for 300 cycles of updating and include data from every trial in which a word was successfully named within 315 cycles.

By examining Figure 1 it is apparent that in the early stages of processing, one of the two meanings of an ambiguous word is slightly more active (this meaning was found to be the more active of the two meanings at the time the phonological units settled) than the meaning of an unambiguous word. The less active meaning of the ambiguous word is at a lower level of activation. The reason for the initial advantage for one meaning of an ambiguous word is that the random starting pattern for the meaning units is, by chance, more likely to be closer to one of the two meanings of an ambiguous word than to the only meaning of an unambiguous word. During later updating cycles, however, the situation reverses and the meaning of an unambiguous word is more highly activated than even the more active meaning of an ambiguous word.

Activation of the target word's phonological pattern grows at virtually the same rate for both ambiguous and unambiguous words. The stronger activation of one meaning of an ambiguous word does not produce an ambiguity advantage in

[4] In a review of an earlier version of this article, J. Rueckl pointed out (personal communication, November 18, 1994) that instead of implementing different sampling rates, faster settling of phonological units ought to occur if the connection weights between orthographic and phonological units were more structured than those between orthographic and meaning units. We think this idea is interesting and highly plausible. In all the simulations reported in this article, however, patterns of activation for words were purely random, and no attempt was made to approximate a structured relationship between orthographic and phonological patterns of words. We hope to explore the consequences of using structured orthographic and phonological patterns in future simulations.

[5] It is possible that the influence of meaning units on phonological units could instead be reduced by introducing structure into the connection weights between orthographic and phonological units, as discussed in footnote 4. The structure inherent in these weights should produce a stronger influence on the phonological units than the unstructured weights between meaning and phonological units. Again, we thank Jay Rueckl for pointing out this possibility.

Table 8
Mean Latencies (in Updating Cycles) and Percentage of Error Rates in Simulation Results

Experiment	Latency U	Latency A	Latency NW	Error U	Error A	Error NW
	Experiment 2					
Word naming						
M^a	290.2	290.8		3.7	3.7	
SD	2.9	2.5		2.6	2.7	
	Experiment 3					
Lexical decision: legal nonwords						
M^b	119.1	116.2	240.0	5.2	3.2	4.9
SD	6.9	5.6	0.0	4.2	3.2	3.2
	Experiment 4					
Lexical decision: illegal nonwords						
M^b	54.5	54.5	80.0	4.0	3.6	5.5
SD	1.9	2.2	0.0	2.1	2.2	2.0
Gaze duration during comprehension						
M^c	629.3	688.8		0.0	0.8	
SD	8.0	17.0		0.1	1.3	

Note. U = unambiguous; A = ambiguous; NW = nonword.
[a]Latency to settle units in the phonological module. [b]Latency to reach criterion energy. [c]Latency to settle units in the meaning module.

naming because the connection weights that link meaning and phonological units are more strongly determined by unambiguous than by ambiguous words. This comes about because of the learning regimen in which each meaning of an ambiguous word is experienced only once; whereas the meaning of an unambiguous words is presented twice. The early activation advantage for ambiguous words, then, seems to be counterbalanced by the stronger influence of unambiguous word meanings on phonological units. We note, however, that the time course of activation of phonological and meaning units may be critical to this balance. If meaning units were sampled at a higher rate, then the point at which activation of meaning among unambiguous words exceeds that of ambiguous words would occur earlier and an advantage for unambiguous words might be observed.

Lexical Decision

The lexical decision results of Experiments 3 and 4 were simulated by using exactly the same model and parameters as in the simulation of naming results. The crucial extension to the model was the assumption that a lexical decision is made on the basis of familiarity invoked by a letter string (e.g., Balota & Chumbley's, 1984, variant of Atkinson & Juola's, 1973, word recognition model; Besner, 1983; Besner & Johnston, 1989). Others have simulated the lexical-decision task by assuming that some set of processing units settle into a stable state. For example, using a model similar to that described here, Joordens and Besner (1994) assumed that lexical decision required full settling of meaning units. Similarly, Kawamoto et al. (1994) and Plaut and Shallice (1993) assumed that lexical decisions required full settling of a set of orthographic units. We have argued elsewhere (Masson & Borowsky, 1995) that full settling, particularly of meaning units, is not necessarily the best characterization of the lexical-decision task.

As an alternative to full settling, there is a feature of Hopfield (1982) networks that suggests itself as a potential means of assessing the familiarity of a stimulus. Recall that the updating procedure used in a Hopfield network can be described as movement into a basin of attraction that corresponds to a learned pattern of activation (a word's representation). A metric known as *energy* can be used to gauge the network's progress down a basin of attraction (Hopfield, 1982). Energy is computed as

$$E = -\sum_{i<j} w_{ij} n_i n_j$$

and monotonically decreases (becomes a larger negative number) as updating cycles continue. Larger negative values for energy occur as pairs of units take on values (± 1) that are consistent with the weight of the connection between them. For example, two units with a positive connection weight make a greater contribution to energy if both units are in the same state. Thus, as the network relaxes into a known state, energy takes on a larger negative value.

To simulate lexical decisions, we computed energy within the orthographic and meaning modules. These measures of energy were intended to reflect the familiarity of a letter string's orthography and meaning. Within the orthographic module, for example, energy was computed by applying the formula for energy to each pair of orthographic units. We did

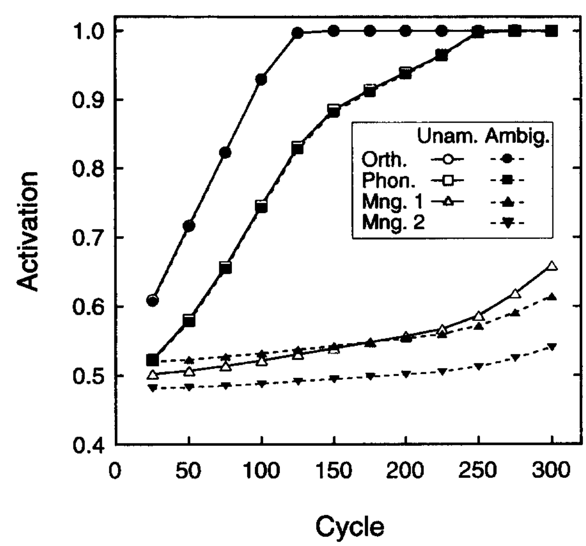

Figure 1. Mean activation of orthographic (Orth.), phonological (Phon.) and meaning patterns (Mng.) of unambiguous (Unam.) and ambiguous (Ambig.) words as a function of update cycle in the distributed memory model. This figure shows the activation for both meanings of ambiguous words, one of which typically is more strongly activated than the other because of the random starting position of the network. Results are shown for all trials in which the phonological units reached a stable state within 315 updating cycles.

not include energy in the phonological module in the simulation under the assumption that orthography and meaning would be the important determinants of a lexical decision. We expect, however, that if energy in the phonological module were included, similar results would be obtained.

During the simulation of a lexical-decision trial, energy was computed after every two updating cycles were run. Energy was not computed after every cycle because that computation is quite time consuming. If computed energy (taken as the sum of energy within the orthographic and meaning modules) reached a minimum criterion before a waiting period expired, then a positive decision was made. Response latency was defined as the number of cycles taken to reach the criterion. If the criterion was not reached by the end of the waiting period, a negative decision was made (see also Anderson, 1976; King & Anderson, 1976). Nonwords were defined by orthographic patterns consisting of random patterns of ±1 across orthographic units. Two nonword patterns were constructed: one was similar to a learned unambiguous word, and the other was similar to a learned ambiguous word.

In preliminary runs of the model, to accomplish two goals, we independently varied the degree of similarity between word and nonword orthographic patterns, energy criterion, and waiting period for simulations of Experiments 3 and 4. First, we sought to obtain an effect of ambiguity in the simulation of Experiment 3, but not in Experiment 4. Second, we aimed to have the model generate error rates that would be roughly similar to those observed in the experiments. The variation of these parameters was conducted with the constraint that the similarity between words and nonwords was greater for Experiment 3 than for Experiment 4. This constraint quite naturally produced the result that accurate lexical decisions could be made after fewer processing cycles in the simulation of Experiment 4.

Satisfactory results were obtained for Experiment 3 when each nonword differed from its similar word in 9 of the 70 orthographic units and for Experiment 4 when the difference between word and nonword pairs was maximal (i.e., they differed in 35 of the 70 orthographic units). These values are reasonable and consistent with the features of the nonwords used in the two experiments (orthographically legal nonwords in Experiment 3 and orthographically illegal nonwords in Experiment 4). The selected energy criteria and waiting periods were −4,600 and 240 cycles for Experiment 3 and −1,200 and 80 cycles for Experiment 4. The shorter waiting period used for the simulation of Experiment 4 was consistent with the much shorter response latencies observed in that experiment, relative to Experiment 3.

A set of 80 runs was performed under each set of parameters to simulate the results of Experiments 3 and 4. The mean correct response latencies measured in cycles to reach the energy criterion and the mean error percentages for each simulation are shown in Table 8. A small but reliable ambiguity advantage in both latency, $F(1, 79) = 13.25$, $MSE = 25.83$, and error percentage, $F(1, 79) = 18.29$, $MSE = 8.42$, was observed in the simulation of Experiment 3, consistent with the empirical data. No ambiguity advantage was found in the simulation of Experiment 4, either in latency ($F < 1$), or in error percentage, $F(1, 79) = 1.47$, $MSE = 4.08$.

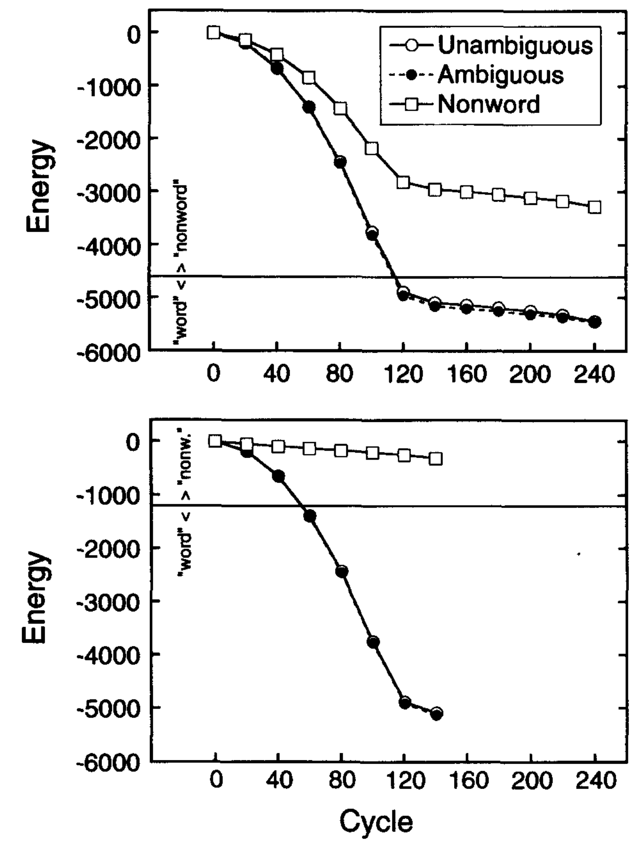

Figure 2. Summed orthographic and meaning energy for the patterns corresponding to unambiguous and ambiguous words and nonwords as a function of update cycle in the distributed memory model. The top section of the figure shows the simulation of Experiment 3, in which pronounceable nonwords were used, and the bottom section shows the simulation of Experiment 4, in which consonant strings as nonwords were used. The horizontal line in each section represents the energy criterion used to make word–nonword decisions. nonw. = nonword.

The network's behavior in simulating the lexical-decision results is summarized in Figure 2, which shows the mean energy value across updating cycles. The values shown in Figure 2 are the sum of energy within the orthographic and the meaning modules. The results for Experiment 3, in which orthographically legal nonwords (i.e., patterns that are similar to those of words) were used, are shown at the top of the figure. A very small advantage can be seen in the energy values for ambiguous words, particularly in the range of 100–200 cycles.

The source of the energy advantage is revealed by inspection of the energy within the orthographic and meaning modules, as shown in Table 9. The top section of the table shows the energy values from the simulation of Experiment 3 for the orthographic and meaning modules separately, as well as the total energy, computed at three different points during processing. These points bracket the average response latency for words. The ambiguous words had a very slight advantage in energy within the orthographic module, but this was just a chance occurrence and was not sustained across repeated simulations.

Table 9
Mean Energy Values for Orthographic and Meaning Modules in the Distributed Memory Model Simulations of the Lexical-Decision Task

	Orthography			Meaning			Total		
Cycle	U	A	NW	U	A	NW	U	A	NW
				Experiment 3					
Lexical decision: legal nonwords									
100	−3,557	−3,584	−1,949	−202	−231	−228	−3,759	−3,815	−2,177
120	−4,657	−4,686	−2,550	−236	−273	−270	−4,893	−4,959	−2,820
140	−4,826	−4,847	−2,641	−269	−313	−313	−5,095	−5,160	−2,954
				Experiment 4					
Lexical decision: illegal nonwords									
40	−561	−562	60	−97	−95	−103	−658	−657	−96
60	−1,269	−1,273	14	−134	−142	−153	−1,403	−1,415	−139
80	−2,265	−2,269	32	−166	−185	−208	−2,431	−2,454	−176

Note. U = unambiguous; A = ambiguous; NW = nonword.

A somewhat larger and more robust energy advantage for ambiguous words was found among the meaning units. These energy values indicated that the meaning units moved into a basin of attraction slightly earlier when an ambiguous rather than an unambiguous word was presented. We believe this phenomenon is a manifestation of the proximity effect described by Joordens and Besner (1994), in which the random starting state of the meaning units is likely to be more similar to one of the two meanings of an ambiguous word than to the single meaning of an unambiguous word.

A rather surprising result is apparent, however, when nonwords are considered. Although the energy values in the orthographic module clearly distinguished words from nonwords, the same cannot be said for the meaning module. It is clear from Table 9 that energy in the meaning module built up at virtually the same rate for nonwords as for words. Thus, although energy differences in the meaning units were responsible for producing the ambiguity advantage in lexical decisions about words, these differences did not contribute to the model's general ability to make word–nonword discriminations.

The buildup of energy in the meaning module during processing of nonwords appears to be a consequence of the meaning units moving into the basin of attraction that was closest to the randomly selected starting pattern for the meaning units. During updating cycles, activation from the orthographic units (representing a nonword pattern unknown to the system) would not be very systematic and therefore would provide very little constraint on the pattern of activation of the meaning units. Thus, the meaning units were, for the most part, free to move into the closest basin of attraction, thereby accumulating energy at a high rate. In our view, this behavior of the meaning units is not realistic and ought to be curbed in some way. For present purposes, however, this problem is not serious because we are primarily interested in the energy values associated with processing ambiguous and unambiguous words.

The lower part of Figure 2 shows the energy functions for the simulation of Experiment 4, in which orthographically illegal nonwords were used. Lexical decisions could be made earlier in that case because the nonword orthographic patterns were unfamiliar and therefore had very small energy values. This observation can be verified by inspecting the lower section of Table 9, which shows the energy values separately for the orthographic and meaning modules. The ability to make word–nonword discriminations earlier means that a decision can be made before the ambiguity advantage in the meaning units has had adequate time to build up. Also note that energy in the meaning units built up more rapidly for nonwords. This occurred because the meaning units were entirely unconstrained by orthographic input in the case of nonwords because of the purely random nature of these orthographic patterns. The meaning units, therefore, were completely free to move into the nearest basin of attraction when a nonword was presented. Once again, we are not seriously concerned about this problem because our major interest lies in the comparison between ambiguous and unambiguous words.

Gaze Duration During Comprehension

In addition to using the distributed memory model to simulate the results of the present naming and lexical-decision experiments, we applied it to a third paradigm. Time spent viewing individual words during reading for comprehension, as measured by fixation duration, can be used to assess word identification processes (e.g., Just & Carpenter, 1980). Rayner and his colleagues (Duffy, Morris, & Rayner, 1988; Rayner & Duffy, 1986; Rayner & Frazier, 1989) measured gaze durations (the sum of the durations of all fixations made on a word before the eye moves to a different word) associated with ambiguous and unambiguous words presented in neutral sentence contexts. The advantage of using a neutral context is that no information is provided to bias participants toward one of the possible meanings of an ambiguous word, making this paradigm similar to the naming and lexical-decision paradigms in that respect.

Gaze duration results show that participants spend significantly more time viewing ambiguous words with equally frequent meanings than either unambiguous words or ambiguous words with one dominant meaning. This ambiguity disadvantage stands in contrast to the ambiguity advantage found in our Experiment 3 with lexical decision. Rayner and Frazier (1989) have argued that the ambiguity disadvantage in gaze duration is a result of accessing all of the common meanings of an ambiguous word, and they suggested that the selection of one of the meanings occurs as a by-product of later integration with a disambiguating context. This argument runs counter to Joordens and Besner's (1994) assumption that participants choose a particular meaning of an ambiguous word when identifying it. Joordens and Besner implicitly make this assumption by their requirement that full settling on a particular meaning must occur in the meaning units for a lexical decision to be made.

We assume that in a comprehension task, unlike lexical-decision and naming tasks, participants must instantiate a substantial portion of a word's meaning. In the simulation results reported here, this assumption was realized by requiring the model to run until the meaning units reached a stable state. Other possibilities exist, such as requiring the meaning units to reach a specified degree of activation (e.g., 70% of the units are in the appropriate state for the target word's meaning), but we have not yet explored them. Gaze duration results were simulated by running the distributed memory model until the meaning units reached a stable state, as defined by an entire pass through the module without any unit changing state. A set of 80 runs was performed by using the same version of the model as in the earlier simulations. If the meaning units failed to settle within 1,000 cycles, the trial was ended and was counted as an error. A relatively large number of cycles were permitted because we sought to minimize errors. The method of obtaining gaze duration data of the type reported by Rayner and his colleagues (e.g., Rayner & Frazier, 1989) does not lend itself to measurement of error rates, and we assume that participants engage a word for sufficient time to permit an accurate reading on nearly all occasions.

The mean cycles required for the meaning units to settle and the mean percentage of error trials are shown in Table 8. There was a strong latency advantage for unambiguous words in this simulation, $F(1, 79) = 707.77$, $MSE = 200.19$, and also a small advantage in percentage of errors, $F(1, 79) = 30.42$, $MSE = 0.86$.

We monitored the activation in all three modules as the model moved toward a stable state in the meaning units. The progress over time in each module is shown in Figure 3, which reveals a number of interesting features of the model's behavior. First, the asymptotic activation of the two meanings of the ambiguous words, on average, was substantially below the maximum of 1.0. This result reflects the fact that on nearly 92% of the trials involving ambiguous words, the meaning units settled into a state in which both meanings were partially activated (one more strongly than the other) but did not correspond to the meaning of any single word. Joordens and Besner (1994) referred to these states as *blend states* and considered them as errors in their simulation of the lexical-decision task. In the analysis of the simulation reported here, however, trials such as these were not considered errors.

Our view is that blend states reflect the partial instantiation

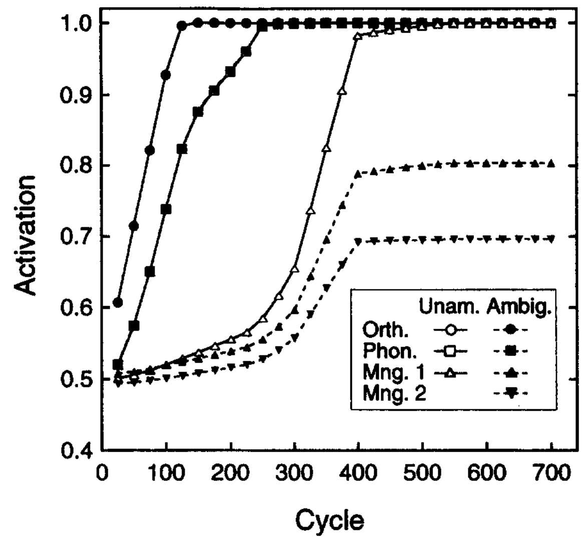

Figure 3. Mean activation of orthographic (Orth.), phonological (Phon.), and meaning patterns (Mng.) of unambiguous (Unam.) and ambiguous (Ambig.) words as a function of update cycle in the distributed memory model. This figure shows the activation for both meanings of ambiguous words, one of which typically is more strongly activated than the other because of the random starting position of the network. Results are shown for all trials in which the meaning units reached a stable state within 1,000 updating cycles.

of two different meanings of an ambiguous word. With no disambiguating contextual information, one meaning typically is more strongly activated on the basis of the random starting pattern of the meaning units. This meaning typically is prevented from achieving full activation, however, because the influence of orthographic units on the meaning units maintains partial activation of the alternative meaning. It might be argued that a blend should be considered an error because it amounts to an instantiation of an incoherent mixture of features (e.g., in the case of the word *buck,* a green piece of paper with antlers). This argument overlooks the natural coherence associated with the subsets of meaning units that instantiate the different meanings of an ambiguous word.

The coherence we have in mind can be seen most easily in the case of a sparse representation of meaning, but we suspect the ideas apply to a full distributed representation as well. In sparsely coded representations, each semantic feature is represented by a single unit, and the meaning of any single word involves activation of only a small subset of these units or features (e.g., Plaut & Shallice, 1993). An ambiguous word would have two (or more) such subsets of features associated with it (one subset for each distinct meaning), and each subset likely would partially be activated when the word's orthographic pattern is presented. Having some members of each subset activated does not mean, however, that an incoherent blend arises. Rather, activated units within each subset cohere with one another by virtue of the relatively strong connection weights between them. Connection weights between units from different subsets would be weaker. Thus, the pattern of activated units and the connection weights within the meaning module provide a basis for segregating the different elements of an ambiguous word's multiple meanings. Although we have not implemented a mechanism for separately accessing different subsets of units related to the different meanings of an ambiguous word, we expect that effective methods could be constructed on the basis of the coherence in these subsets.

The functions in Figure 3 also indicate that the growth of activation of meaning for unambiguous words and for the more active of the two meanings of ambiguous words cross each other at an earlier point than was the case for the simulation of naming data, shown in Figure 1. Moreover, the initial discrepancy between the two ambiguous word meanings is smaller in Figure 3. These differences appear to be due to a selection artifact. The simulation of naming data produced a greater proportion of errors (failure to settle the phonological units before a deadline) than the simulation of gaze duration data. The typical naming error trial appears to have occurred when the initial state of the meaning units gave little or no activation advantage to one of the two meanings of an ambiguous word.

Finally, it is clear that the advantage in accrual of activation associated with unambiguous words holds even before the meaning units reach asymptotic activation. Therefore, other methods of defining gaze durations that might be based on activation of meaning units would also yield an advantage for unambiguous words.

Discussion

The most impressive aspect of the simulation results is the fact that the distributed memory model was able to produce the pattern of ambiguity effects across three different paradigms, consisting of an ambiguity advantage (lexical decision), an ambiguity disadvantage (gaze duration during comprehension), and no effect of ambiguity (naming). The simulation of these effects was achieved by adopting plausible criteria for defining the response on each task while holding basic parameters of the model constant. The model's ability to accommodate all three ambiguity effects might be interpreted as an example of a connectionist model that is too powerful to be meaningful, inasmuch as it is capable of generating any of the three possible ambiguity effects (Massaro, 1988). We contend, however, that the different criteria used to obtain simulations of the different tasks and the behavior of the model under these criteria reveal some useful principles regarding semantic ambiguity.

The learning regimen we used was defined such that the presentation of an ambiguous word incorporated only one meaning of the word. Consequently, the meanings of ambiguous words in comparison to those of unambiguous words were necessarily more weakly represented both in the connection weights within the meaning module itself and in the connection weights between the meaning module and the orthographic and phonological modules. The differential representation of ambiguous and unambiguous word meanings in the connection weights leads to an inherent advantage for unambiguous words, as revealed in the simulation of gaze duration; the stable states associated with the meaning of unambiguous words were reached in fewer updating cycles.

There is also a source of advantage for ambiguous words, however, that derives from the proximity effect pointed out by Joordens and Besner (1994). For an ambiguous word, two different patterns of activation in the meaning units are associated with a single orthographic pattern. When a word is presented for identification, the meaning units begin in a random pattern. On average, a random pattern of activation in the meaning units is more likely to be close to one of the two possible meanings of an ambiguous word than to the single meaning of an unambiguous word. It is this initial advantage that, in the distributed memory model, has the potential to generate an ambiguity advantage such as that observed in the lexical-decision task.

Despite the success of the distributed memory model's simulation of ambiguity effects across three word identification paradigms, there were some aspects of the results that could be considered troublesome. Of primary concern is the fact that the relative amount of time taken by the model to perform the three different tasks, as measured in updating cycles, did not conform to the ordering of actual time to perform the tasks. Empirical data typically show that gaze durations are shorter than naming latencies, which are shorter than lexical-decision latencies. An examination of Table 8 reveals that the model produced the reverse ordering of tasks. Only in the case of the two variants of the lexical-decision task were the overall task latencies ordered correctly by the model.

We do not see this situation as fatal for the model for two reasons. First, the model is intended to capture only some of the processes required to complete the tasks in question. Other processes, not included in the model, would require additional time. In the case of lexical decision, for example, once the computed energy of the orthographic and meaning modules exceeds the criterion, a manual response must be prepared and executed. For the naming task, a vocal response must be initiated. In the case of gaze durations, the next eye movement must be made. We know very little about the time required for these different task components. Second, we have only intuition to guide us with respect to translating updating cycles into a measure of real time. We have assumed, for example, that updating cycles within each module takes the same amount of time and that the function relating updating cycles to real time is constant across tasks. Neither assumption is necessarily correct. For example, under a general capacity model (e.g., Just & Carpenter, 1992), it is possible that updating cycles might take longer when task demands are heavier. In the case of lexical decision, for example, it is assumed that participants must evaluate the familiarity of a stimulus. This evaluation is realized in the model by a time consuming computation of energy. In fact, although the model generated lexical decisions in fewer cycles than it required to produce a naming response, it actually took longer to compute the information needed to perform the former task. For these reasons, although we are confident in making within-task comparisons of latency results, we are much more tentative regarding arguments about cross-task comparisons of latency.

A second concern about the model stems from the simulation of lexical decisions. We based decisions on the sum of energy in the orthographic and meaning modules, as shown in Figure 2. Energy in the orthographic module reached a much larger (negative) asymptotic value for words than for nonwords and was responsible for the model's ability to discriminate between words and nonwords (see Table 9). The energy in the meaning module, however, accrued at a similar rate for words and nonwords. That is, despite presenting the model with an unknown orthographic pattern (even a purely random pattern in the case of the simulation of Experiment 4), the meaning units still moved into a basin of attraction, and energy among those units changed accordingly. The activation sent to meaning units by orthographic units in the case of nonwords is noisy and therefore exerts less influence on the meaning units than is the case for words. Unconstrained by a coherent orthographic or phonological message, the meaning units moved into the basin of attraction that was closest to the arbitrary starting pattern of those units. This situation would be akin to a participant thinking of the first word meaning that came to mind while viewing a nonword letter string. Our intuition is that although participants may engage in such behavior, it is unlikely to happen as early in the course of a lexical-decision trial as it did in the model. To prevent the model from engaging in this behavior, it might be necessary to allow activity in the orthographic and phonological units to influence the rate at which updating in the meaning units takes place. For example, the updating rate for meaning units might initially be very low but would increase as the energy in the orthographic or phonological modules grows, indicating that a potentially familiar stimulus has been presented.

Although the details of the distributed memory model we used to generate the simulations reported here are unlikely to be entirely correct, we see merit in adopting as working hypotheses two ideas that are embodied in the model. First, the advantage of ambiguity springs from a short-lived proximity advantage in the meaning units. Second, the disadvantage of ambiguity results from competition between the two meanings of an ambiguous word that are invoked by their shared orthographic and phonological patterns. This competition makes activation of one meaning of an ambiguous word less efficient, and even less likely, than in the case of an unambiguous word.

General Discussion

In Experiments 1 and 2 we examined the effects of semantic ambiguity (i.e., whether a word has one or more than one meaning) in a standard naming task. Although an advantage for naming ambiguous words has been reported by Balota et al. (1991), Fera et al. (1992), and Hino and Lupker (1993), we found no reliable advantage with either the Fera et al. items (Experiment 1) or our own (Experiment 2). Turning to the lexical-decision task in Experiment 3, we found a significant ambiguity advantage, which served to replicate similar results reported by other researchers (e.g., Jastrzembski, 1981; Jastrzembski & Stanners, 1975; Kellas et al., 1988; Millis & Button, 1989; Pugh et al., 1994; Rubenstein et al., 1970, 1971). In Experiment 4, we tested the Kawamoto et al. (1994) assumption that the locus of the ambiguity advantage in lexical decision is at the orthographic level of processing. We used orthographically illegal nonwords that could be more quickly and easily discriminated from the word stimuli, thus encouraging a more rapid and, arguably, a less conceptually influenced lexical decision. Contrary to the Kawamoto et al. assumption, there was no reliable ambiguity advantage under these circumstances.

A necessary caveat, however, concerns the robustness of the semantic ambiguity advantage observed in the lexical-decision task. Despite careful, item-by-item matching of ambiguous and unambiguous words on a variety of word characteristics, the potential remains for incorrectly concluding that an ambiguity advantage exists in lexical decision and similarly for the null effect of ambiguity in naming. This is so because we cannot be exhaustive with respect to the variables on which the items are matched. Previously unidentified variables that affect word identification performance will continue to appear in the literature, potentially compromising the robustness of any between-stimuli effect (e.g., word frequency, semantic or phonological ambiguity, neighborhood density, and so forth). For example, had we not collected subjective familiarity ratings (following Gernsbacher's, 1984, demonstration of a confound between familiarity and ambiguity) and considered them for inclusion in our regression analyses, we would have concluded that there is no reliable unique effect of ambiguity (by items) on response latency in the lexical-decision task. Thus, it remains to be seen whether the ambiguity advantage that we observed in the lexical-decision task (and, for that matter, the lack of an ambiguity effect in naming) will prevail when future word identification variables are taken into consideration.

A simulation of the naming task used in Experiments 1 and 2 demonstrated the ability of the distributed memory model to simulate the empirical observation of no ambiguity advantage in word naming. The same model successfully simulated the lexical-decision task and the empirical observation of an ambiguity advantage in this task by tracking how close the network is to a known or familiar state (i.e., energy) when presented a letter string and by imposing a criterion that served to distinguish words from nonwords. The distributed memory model also captured the empirical observation of no reliable ambiguity advantage when orthographically illegal nonwords are used. Finally, the ambiguity disadvantage on gaze duration that is seen when participants are asked to read for comprehension (Duffy et al., 1988; Rayner & Duffy, 1986; Rayner & Frazier, 1989) was also accommodated by this model. Thus, a version of the distributed memory model common to all of these simulations was successful in accommodating (a) the absence of an ambiguity effect in naming performance, (b) the ambiguity advantage seen in lexical-decision performance, and (c) the ambiguity disadvantage seen in gaze duration during reading for comprehension. We acknowledge, however, that modeling semantic ambiguity effects is a dynamic process itself. As we have demonstrated in this article, there is a high degree of flexibility in distributed models that will permit one to accommodate an increasingly complicated set of empirical findings. We are not espousing the distributed memory model here as the final model of semantic ambiguity effects. Instead, we have provided a general solution in response to a general problem that was pointed out by Joordens and Besner (1994) and have extended the model to accommodate existing semantic ambiguity effects.

The distributed memory model also generates some interesting, and testable, predictions about semantic ambiguity. In general, the model predicts a potential ambiguity advantage whenever the task requires a judgment that can be made on the basis of familiarity with the stimulus (e.g., the lexical-decision task used in Experiment 3 and perhaps the "semantic access" task used by Pugh et al., 1994, in which participants were asked to press a button when any meaning of the target word has come to mind). However, the distributed memory model predicts an ambiguity disadvantage whenever the task requires settling on a particular meaning of a word (e.g., the studies reviewed earlier involving gaze duration during reading for comprehension) or on a subset of a word's semantic features. For example, a classification task requiring living–nonliving judgments would be simulated by requiring an appropriate subset of the meaning units to settle. This settling would be achieved sooner in the case of an unambiguous word (just as full settling occurs sooner for an unambiguous word). It remains for future research to determine whether this prediction will be borne out by the data.

An alternative account of the present data concerns the similar null effects of ambiguity in both the naming task and the lexical-decision task with orthographically illegal nonwords. It could be argued that both of these tasks yielded null effects of ambiguity because participants did not have to carry out conceptual processing and that both tasks could be simulated without any contribution of conceptual processing. Although this account is parsimonious in that it appeals to a common process to accommodate the absence of ambiguity effects in these two tasks, it fails to account for other empirical observations in both of these tasks. First, Joordens and Besner (1992) have reported a small, but significant, effect of semantic priming over an intervening word in a rapid, continuous naming task. If there were no conceptual processing during the naming task, then such an effect could not exist. Furthermore, it is difficult to account for this priming effect with a nonsemantic mechanism, say, by direct associative links between representations at a subsemantic level because of the fact that this priming effect occurs over an unrelated intervening word. Second, in our Experiment 4 data (lexical decision with illegal nonwords), we found a significant, albeit reduced, correlation between response latency and word frequency, a variable that has been argued to reflect semantic-level processing in the lexical-decision task (Borowsky & Besner, 1993). Thus, it appears that at least some degree of conceptual processing was occurring during both of these tasks, and our simulations reflect this view.

Another model that appeared to be promising with respect to simulation of the ambiguity advantage in lexical decision and the ambiguity disadvantage in gaze duration during reading for comprehension was that of Kawamoto et al. (1994). However, how would this model handle the lack of any effect of ambiguity in naming? A logical extension of the Kawamoto et al. model to handle the naming task would incorporate a set of phonological units. Word naming would be simulated on the basis of activation in the phonological units (Kawamoto, 1993). Because the phonology of a semantically ambiguous word is consistent regardless of its multiple meanings, although the connections between phonology and meaning are not, the error–correction learning algorithm will serve to compensate for inconsistency by increasing the influence from consistent sources. Thus, the weights connecting phonological units will be more influenced by the pattern of activation for an ambiguous word than for an unambiguous word. In fact, the Kawamoto et al. model makes a straightforward prediction regarding performance on any task that requires a consistent response to be made to an ambiguous word (e.g., naming and lexical decision). Such tasks will yield a processing advantage for ambiguous words at the level where the consistency occurs: the phonological level during the naming task and the orthographic level during the lexical-decision task. As long as the level that is monitored for the response is the level where the consistency occurs, an ambiguity advantage in performance will result. On the other hand, if the level that is monitored for the response is, instead, where an inconsistency in processing occurs (e.g., the conceptual level during reading for comprehension), an ambiguity disadvantage in performance will result. Nonetheless, if we are correct in concluding that there is no ambiguity processing advantage during word naming, then it is clear that the Kawamoto et al. model does not adequately capture the full range of ambiguity effects discussed in this article.

One solution to the problem inherent in assessing the between-stimuli effect of semantic ambiguity is to create an analogous within-stimuli manipulation of semantic ambiguity. If an orthographically legal and pronounceable nonword (e.g., *grick* is considered) as a potential noun that can arbitrarily be mapped to one concept (unambiguous) or multiple concepts (ambiguous), then a manipulation of semantic ambiguity within-stimuli is conceivable.[6] For example, in a learning phase a participant could be asked to consider the concept of broken glass and to associate the orthographic pattern *grick* to this concept. A researcher could be reasonably certain that *grick* would unambiguously represent the concept of broken glass for that participant. On the other hand, if the participant is later asked to consider another concept, say, a pile of leaves, and is also asked to associate that concept also to the orthographic pattern *grick*, a researcher could be reasonably certain that *grick* would now ambiguously represent both the concept of broken glass and the concept of a pile of leaves. After the learning phase, the participant's word identification performance could then be tested in any one of the tasks that have been discussed in this article. If our conclusions from the present experiments are correct, then the same pattern of word identification performance should also be obtained in this paradigm.

[6] We note that other researchers have had ideas along this line. D. Besner and S. Joordens (personal communication, November 18, 1994) have proposed a similar experiment. Also, Rueckl and Olds (1993) have conducted a similar learning study but did not examine lexical decision or naming performance. They did find, however, that there was no ambiguity effect on identification accuracy (response latency was not a dependent measure in their experiments).

References

Anderson, J. R. (1976). *Language, memory, and thought.* Hillsdale, NJ: Erlbaum.

Andrews, S. (1989). Frequency and neighborhood effects on lexical access: Activation or search? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 802–814.

Andrews, S. (1992). Frequency and neighborhood effects on lexical access: Lexical similarity or orthographic redundancy? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 234–254.

Atkinson, R. C., & Juola, J. F. (1973). Factors influencing speed and accuracy of word recognition. In S. Kornblum (Ed.), *Attention and performance IV* (pp. 583–612). New York: Academic Press.

Balota, D. A., & Chumbley, J. I. (1984). Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. *Journal of Experimental Psychology: Human Perception and Performance, 10,* 340–357.

Balota, D. A., & Chumbley, J. I. (1985). The locus of word-frequency effects in the pronunciation task: Lexical access and/or production? *Journal of Memory and Language, 24,* 89–106.

Balota, D. A., Ferraro, F. R., & Conner, L. T. (1991). On the early influence of meaning in word recognition: A review of the literature. In P. J. Schwanenflugel (Ed.), *The psychology of word meanings* (pp. 187–222). Hillsdale, NJ: Erlbaum.

Besner, D. (1983). Basic decoding components in reading: Two dissociable feature extraction processes. *Canadian Journal of Psychology, 37,* 429–438.

Besner, D., & Johnston, J. C. (1989). Reading and the mental lexicon: On the uptake of visual information. In W. Marslen-Wilson (Ed.), *Lexical representation and process* (pp. 291–316). Cambridge, MA: MIT Press.

Borowsky, R., & Besner, D. (1993). Visual word recognition: A multistage activation model. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 813–840.

Burt, J. S., & Humphreys, M. S. (1993). Delayed priming of the pronunciation of inconsistent words and pseudowords. *Journal of Memory and Language, 32,* 743–765.

Clark, H. H. (1973). The language-as-a-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior, 12,* 335–359.

Coltheart, M., Davelaar, E., Jonasson, J. T., & Besner, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.), *Attention and performance VI* (pp. 535–555). Hillsdale, NJ: Erlbaum.

Duffy, S. A., Morris, R. K., & Rayner, K. (1988). Lexical ambiguity and fixation times in reading. *Journal of Memory and Language, 27,* 429–446.

Fera, P., Joordens, S., Balota, D. A., Ferraro, F. R., & Besner, D. (1992, November). *Ambiguity in meaning and phonology: Effects on naming.* Paper presented at the 33rd annual meeting of the Psychonomic Society, St. Louis, MO.

Ferraro, R. F., & Kellas, G. (1990). Normative data for number of word meanings. *Behavior Research Methods, Instruments, & Computers, 22,* 491–498.

Forster, K. I., & Bednall, E. S. (1976). Terminating and exhaustive search in lexical access. *Memory & Cognition, 4,* 53–61.

Gernsbacher, M. A. (1984). Resolving 20 years of inconsistent interactions between lexical familiarity and orthography, concreteness, and polysemy. *Journal of Experimental Psychology: General, 113,* 256–281.

Grainger, J. (1990). Word frequency and neighborhood frequency effects in lexical decision and naming. *Journal of Memory and Language, 29,* 228–244.

Hebb, D. O. (1949). *The organization of behavior.* New York: Wiley.

Hino, Y., & Lupker, S. J. (1993, July). *Polysemy effects: Evidence for dual access routes to word meanings.* Paper presented at the third annual meeting of the Canadian Society for Brain, Behavior, and Cognitive Science, Toronto, Ontario, Canada.

Hinton, G. E., & Shallice, T. (1991). Lesioning an attractor network: Investigations of acquired dyslexia. *Psychological Review, 98,* 74–95.

Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Science, 79,* 2554–2558.

Jastrzembski, J. E. (1981). Multiple meanings, number of related meanings, frequency of occurrence, and the lexicon. *Cognitive Psychology, 13,* 278–305.

Jastrzembski, J. E., & Stanners, R. F. (1975). Multiple word meanings and lexical search speed. *Journal of Verbal Learning and Verbal Behavior, 14,* 534–537.

Joordens, S., & Besner, D. (1992). Priming effects that span an intervening unrelated word: Implications for models of memory representation and retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 483–491.

Joordens, S., & Besner, D. (1994). When banking on meaning is not (yet) money in the bank: Explorations in connectionist modeling. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 1051–1062.

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review, 87,* 329–354.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review, 99,* 122–149.

Kawamoto, A. H. (1993). Nonlinear dynamics in the resolution of lexical ambiguity: A parallel distributed processing account. *Journal of Memory and Language, 32,* 474–516.

Kawamoto, A. H., Farrar, W. T., & Kello, C. (1994). When two meanings are better than one: Modeling the ambiguity advantage using a recurrent distributed network. *Journal of Experimental Psychology: Human Perception and Performance, 20,* 1233–1247.

Kellas, G., Ferraro, R. F., & Simpson, G. B. (1988). Lexical ambiguity and the time-course of attentional allocation in word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 14,* 601–609.

King, D. R. W., & Anderson, J. R. (1976). Long-term memory search: An intersecting activation process. *Journal of Verbal Learning and Verbal Behavior, 15,* 587–606.

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

Lukatela, G., & Turvey, M. T. (1994). Visual lexical access is initially phonological: 2. Evidence from phonological priming by homophones and pseudohomophones. *Journal of Experimental Psychology: General, 123,* 331–353.

Massaro, D. W. (1988). Some criticisms of connectionist models of human performance. *Journal of Memory and Language, 27,* 213–234.

Masson, M. E. J. (1991). A distributed memory model of context effects in word identification. In D. Besner & G. W. Humphreys (Eds.), *Basic processes in reading: Visual word recognition* (pp. 233–263). Hillsdale, NJ: Erlbaum.

Masson, M. E. J. (1995). A distributed memory model of semantic priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 3–23.

Masson, M. E. J., & Borowsky, R. (1995). Unsettling questions about semantic ambiguity in connectionist models: Comment on Joordens and Besner (1994). *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 509–514.

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

Millis, M. L., & Button, S. B. (1989). The effect of polysemy on lexical decision time: Now you see it, now you don't. *Memory & Cognition, 17,* 141–147.

Monsell, S., Doyle, M., & Haggard, P. (1989). Effects of frequency on visual word recognition tasks: Where are they? *Journal of Experimental Psychology: General, 118,* 43–71.

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

Perfetti, C. A., & Bell, L. (1991). Phonemic activation during the first 40 ms of word identification: Evidence from backward masking and masked priming. *Journal of Memory and Language, 30,* 473–485.

Perfetti, C. A., Bell, L., & Delaney, S. (1988). Automatic phonetic activation in silent word reading: Evidence from backward masking. *Journal of Memory and Language, 27,* 59–70.

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

Pugh, K. R., Rexer, K., & Katz, L. (1994). *Effects of orthographic neighborhood distribution in visual word recognition.* Manuscript submitted for publication.

Random House dictionary. (1978). New York: Ballantine.

Rayner, K., & Duffy, S. A. (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory & Cognition, 14,* 191–201.

Rayner, K., & Frazier, L. (1989). Selection mechanisms in reading lexically ambiguous words. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 15,* 779–790.

Rubenstein, H., Garfield, L., & Millikan, J. A. (1970). Homographic entries in the internal lexicon. *Journal of Verbal Learning and Verbal Behavior, 9,* 487–492.

Rubenstein, H., Lewis, S. S., & Rubenstein, M. A. (1971). Homographic entries in the internal lexicon: Effects of systematicity and relative frequency of meanings. *Journal of Verbal Learning and Verbal Behavior, 10,* 57–62.

Rueckl, J. G. (1995). Ambiguity and connectionist networks: Still settling into a solution—Commentary on Joordens and Besner (1994). *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 501–508.

Rueckl, J. G., & Olds, E. M. (1993). When pseudowords acquire meaning: Effect of semantic associations on pseudoword repetition priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 515–527.

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

Shulman, H. G., Hornak, R., & Sanders, E. (1978). The effects of graphemic, phonetic, and semantic relationships on access to lexical structures. *Memory & Cognition, 6,* 115–123.

Van Orden, G. C., Johnston, J. C., & Hale, B. L. (1988). Word identification in reading proceeds from spelling to sound to meaning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 371–385.

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

Appendix

Characteristics of Ambiguous (A) and Unambiguous (U) Words Used in Experiment 2

Word		Orthographic error score		Phonological error score		Word frequency		Familiarity rating		Word length		Orthographic neighborhood density		No. of higher frequency neighbors	
A	U	A	U	A	U	A	U	A	U	A	U	A	U	A	U
ace	ate	4.76	4.93	6.90	5.70	15	16	4.63	6.63	3	3	12	10	5	2
bat	bet	3.76	5.10	3.72	2.89	18	20	5.05	5.05	3	3	23	22	13	12
bluff	bloat	8.64	8.60	4.65	5.75	8	8	3.47	3.95	5	5	2	4	0	0
bolt	bite	10.19	6.76	4.83	3.22	10	10	3.95	5.74	4	4	10	15	0	5
boot	buzz	5.18	11.51	3.07	2.77	13	13	5.84	4.42	4	4	12	14	2	7
bowl	bend	6.92	4.53	3.58	4.81	23	24	6.16	5.26	4	4	5	2	0	1
buck	bold	5.08	3.71	3.54	2.75	20	21	4.58	4.53	4	4	17	11	3	1
charge	chance	6.04	3.88	3.85	3.12	122	131	5.11	5.84	6	6	12	5	3	0
chest	chain	5.23	3.52	4.42	2.74	53	50	5.58	4.79	5	5	6	6	0	4
coast	curve	4.44	3.97	3.66	2.59	61	45	5.53	5.26	5	5	10	7	2	3
dash	duke	8.17	11.26	4.84	6.54	11	11	3.84	2.95	4	4	20	13	4	3
date	deep	6.09	5.51	2.79	3.09	103	109	5.89	6.11	4	4	16	9	8	3
deck	disk	4.05	4.95	2.83	3.27	23	25	4.79	5.53	4	4	19	17	5	7
draft	drain	5.94	4.51	3.98	3.27	24	18	4.58	5.00	5	5	4	9	1	0
drag	dish	3.52	5.05	3.04	3.81	15	16	4.11	6.37	4	4	16	6	5	2
duck	dusk	5.20	8.38	3.26	3.74	9	9	5.00	4.37	4	4	8	15	2	5
field	force	—	3.42	—	2.80	274	230	5.21	5.00	5	5	15	16	5	7
fine	fear	4.34	3.83	3.03	3.17	161	127	6.00	6.21	4	4	12	14	0	1
foil	fame	6.73	5.01	3.61	2.96	20	18	3.53	4.84	4	4	8	4	4	1
grade	grown	4.67	5.30	2.73	3.16	35	43	6.32	5.26	5	5	6	5	4	0
grave	grain	5.41	4.25	3.69	3.51	33	27	4.74	4.74	5	5	9	20	1	6
hail	hunt	4.34	5.46	3.73	4.65	10	10	3.74	4.74	4	4	17	6	4	1
hog	hug	7.64	6.68	7.55	4.54	3	3	3.89	6.00	3	3	17	9	5	2

Appendix (*continued*)

Word		Orthographic error score		Phonological error score		Word frequency		Familiarity rating		Word length		Orthographic neighborhood density		No. of higher frequency neighbors	
A	U	A	U	A	U	A	U	A	U	A	U	A	U	A	U
hound	hitch	7.57	7.96	5.69	4.77	7	5	3.53	3.53	5	5	9	9	1	3
jerk	jolt	22.99	15.80	7.66	9.74	2	4	5.32	3.37	4	4	16	7	3	0
joint	jump	11.46	4.92	5.56	3.17	39	24	4.63	6.42	5	4	5	6	3	2
land	love	3.77	4.57	3.27	2.61	217	232	6.00	6.84	4	4	9	4	0	0
lean	loud	4.98	8.46	5.28	5.38	20	20	4.37	6.47	4	4	6	4	0	2
light	large	3.38	4.80	2.64	3.64	333	361	6.63	6.63	5	5	13	5	6	1
loaf	lint	19.11	4.68	5.03	3.34	4	4	5.05	4.21	4	4	18	9	3	2
match	meat	4.35	3.89	2.72	2.87	41	45	5.42	6.26	5	4	16	2	5	2
mint	maze	6.01	5.79	3.04	2.85	7	6	5.47	4.05	4	4	21	2	4	0
miss	mind	4.16	3.92	2.53	2.82	258	325	5.58	6.16	4	4	2	3	1	1
palm	plug	6.20	4.75	8.23	3.62	22	23	4.32	5.26	4	4	9	7	2	2
pet	pig	8.47	6.41	3.93	3.53	8	8	5.84	5.47	3	3	12	10	3	4
pound	pond	7.75	7.33	4.12	3.39	28	25	5.42	4.68	5	4	4	5	0	0
punch	plumb	7.65	24.18	5.25	4.13	5	5	4.89	3.21	5	5	13	17	3	6
ram	rum	5.21	8.34	2.87	3.14	2	3	3.26	4.79	3	3	10	9	2	1
rock	rain	4.70	3.62	2.79	2.75	75	70	6.21	6.89	4	4	10	6	2	2
roll	rice	3.39	5.18	2.85	2.94	35	33	5.37	5.74	4	4	8	9	7	3
row	rev	4.30	10.06	5.63	4.00	35	33	5.84	2.32	3	3	2	8	0	2
sack	sane	6.73	6.57	4.06	3.65	8	8	4.47	4.79	4	4	10	4	3	0
screen	stream	6.69	4.00	2.55	5.05	48	51	4.84	4.68	6	6	6	15	3	8
seal	soup	4.17	11.74	3.15	6.20	17	16	4.26	6.00	4	4	9	4	2	1
shed	shoe	8.49	5.77	3.60	8.27	11	14	4.74	6.79	4	4	22	17	15	7
sink	soap	3.97	5.95	3.16	4.51	23	22	5.79	6.32	4	4	17	15	9	9
spade	steak	11.57	5.01	3.00	9.75	10	10	2.89	5.63	5	5	14	22	7	10
spring	spent	5.64	4.87	3.27	4.00	127	104	5.68	5.63	6	5	3	3	0	0
stall	steam	5.87	4.21	6.71	4.67	18	17	4.00	5.16	5	5	8	4	3	0
star	stuck	4.49	6.26	3.66	3.28	25	23	5.74	4.95	4	5	9	18	1	4
steer	stack	8.22	4.43	3.48	3.95	9	9	4.47	4.37	5	5	6	11	3	3
stick	stuff	8.20	6.00	4.68	4.19	39	32	5.42	5.89	5	5	10	5	1	1
strike	strife	4.96	7.73	3.5	5.04	6	6	4.89	3.05	6	6	9	7	2	2
strip	storm	7.53	5.35	3.78	4.20	30	26	4.68	6.05	5	5	4	6	1	2
swamp	sworn	9.94	12.94	9.46	6.03	5	5	3.89	4.32	5	5	5	4	0	1
switch	swept	5.13	4.71	3.05	2.88	43	34	4.84	4.58	6	5	25	20	6	10
tag	tee	5.17	4.67	3.24	5.20	5	5	4.74	2.89	3	3	17	17	3	8
tick	turf	7.65	18.51	3.46	5.40	3	3	3.58	3.53	4	4	6	9	0	1
tip	tea	3.72	5.69	3.51	5.75	22	28	5.58	6.32	3	3	4	7	1	1
tire	tent	4.52	5.49	3.28	2.97	22	20	5.63	5.37	4	4	9	4	0	1
type	town	6.43	3.33	2.87	2.63	200	212	5.79	5.89	4	4	4	5	0	0
vault	valve	24.19	9.95	8.34	7.33	2	3	3.32	3.89	5	5	12	16	1	0
wake	wash	3.73	4.59	2.91	3.15	23	37	5.79	6.47	4	4	14	12	0	3
watch	worth	3.54	3.26	2.82	2.75	81	94	6.11	5.32	5	5	17	13	6	1
Mean		6.70	6.50	4.09	4.13	46	46	4.94	5.14	4.4	4.3	10.9	9.4	2.9	2.8

Note. Dashes indicate error scores were not available.

Received August 17, 1994
Revision received December 13, 1994
Accepted January 19, 1995 ■