Dynamics of activation of semantically similar concepts during spoken word recognition

DANIEL MIRMAN
Moss Rehabilitation Research Institute, Philadelphia, Pennsylvania

AND

JAMES S. MAGNUSON
University of Connecticut, Storrs, Connecticut
and Haskins Laboratories, New Haven, Connecticut

Semantic similarity effects provide critical insight into the organization of semantic knowledge and the nature of semantic processing. In the present study, we examined the dynamics of semantic similarity effects by using the visual world eyetracking paradigm. Four objects were shown on a computer monitor, and participants were instructed to click on a named object, during which time their gaze position was recorded. The likelihood of fixating competitor objects was predicted by the degree of semantic similarity to the target concept. We found reliable, graded competition that depended on degree of target–competitor similarity, even for distantly related items for which priming has not been found in previous priming studies. Time course measures revealed a consistently earlier fixation peak for near semantic neighbors relative to targets. Computational investigations with an attractor dynamical model, a spreading activation model, and a decision model revealed that a combination of excitatory and inhibitory mechanisms is required to obtain such peak timing, providing new constraints on models of semantic processing.

In typical speech contexts, listeners must correctly identify and interpret from 100 to 150 words per minute.[1] This is done seemingly without effort, despite the noise and ambiguity inherent in the speech signal and the complexity of the semantic knowledge that must be accessed. Because the process is so fast and the semantic structure is so complex, the dynamics of spoken word recognition are challenging to study. The many different theories regarding the structure of semantic knowledge can be grouped according to a few critical distinguishing properties. One distinction is the granularity of representations, with approaches varying from those in which *concept* is the lowest level of analysis or representation to those in which subconceptual elements or features are the lowest level. In network models of knowledge, this is a distinction between localist and distributed representations. Under the localist view, each concept is a unique node in a network (e.g., Collins & Loftus, 1975; Steyvers & Tenenbaum, 2005) and the connections among the nodes in the network explicitly determine their effects on one another. Under the distributed view, concepts are represented by patterns of activation over the same set of units (e.g., Landauer & Dumais, 1997; Lund & Burgess, 1996; McRae, Cree, Seidenberg, & McNorgan, 2005; Vigliocco, Vinson, Lewis, & Garrett, 2004) and effects of concepts on one another are an emergent property of processing dynamics and the patterns of overlap.

A second critical distinguishing property is the proposed structure of semantic relations. Conceptual relatedness has been hypothesized to depend on membership in the same category (e.g., Chiarello, Burgess, Richards, & Pollock, 1990; Hines, Czerwinski, Sawyer, & Dwyer, 1986), association by co-occurrence in text or speech (e.g., Nelson, McEvoy, & Schreiber, 2004), or shared perceptual, action, or other features (e.g., Barsalou, 1999; McRae et al., 2005; Vigliocco et al., 2004). Of particular interest are cases in which these approaches make different behavioral predictions. There is general agreement that, as a word is processed, words with related meanings are partially activated, but the different approaches make different claims about which meanings are related. Specifically, under a strict category hierarchy view, only category coordinates should be activated; under an association-based view, only associates should be activated; and under a feature-based view, co-activation is determined by feature overlap (although activation can also result from semantic association).

This issue has been addressed in a number of studies using semantic priming, but with mixed results. Shelton and Martin (1992) found priming for associated word pairs but not for semantically related word pairs that were not associated, suggesting that associations—not feature-based semantic relatedness—form the basis of semantic structure. McRae and Boisvert (1998) showed priming for high semantic similarity pairs that were not associated and argued that Shelton and Martin failed to find priming because of the low semantic similarity between primes and targets in their study. Similarly, Cree, McRae, and McNor-

---

D. Mirman, mirmand@einstein.edu

gan (1999) found that priming was determined by feature-based semantic similarity rather than by shared category membership. The results of these last two studies suggest that semantic features form the basis of semantic structure. However, although attractor network simulations by Cree et al. predicted that low levels of feature overlap should still produce priming effects (albeit very small ones), all three failed to find semantic priming for distantly related concepts. What could cause this consistent null result? Is the attractor network sensitive to behaviorally inconsequential degrees of overlap? Or, might the priming paradigm be insufficiently sensitive to detect such differences?

For over 30 years, semantic priming has served as the primary method for studying the time course of word comprehension, and results from priming studies have been central to theoretical development in the field. However, the paradigm has two weaknesses. First, the priming paradigm requires participants to make metalinguistic judgments (such as lexical decision or semantic categorization), which may promote use of strategies that obscure the underlying word-processing dynamics. Second, the time course of word processing must be inferred from experimenter-determined prime durations and from delays between prime and target presentation, which may give an incomplete picture of the time course of word recognition.

The domain of phonological processing has provided a concrete example in which experiments using the priming paradigm have failed to detect small similarity effects. Spoken word recognition studies consistently detect phonological priming for words that share onsets (e.g., *beaker–beetle*), but not for words that share offsets (e.g., *beaker–speaker*; Marslen-Wilson & Zwitserlood, 1989). This pattern is consistent with the hypothesis that only words that match at onset are activated during spoken word recognition, which was taken as support for one class of model of spoken word recognition (e.g., the cohort model; Marslen-Wilson, 1987) and as a challenge to other models (e.g., the TRACE model of speech perception, McClelland & Elman, 1986, and the neighborhood activation model, P. A. Luce & Pisoni, 1998). The empirical picture of competition during spoken word recognition changed when the visual world eyetracking paradigm (VWP; Cooper, 1974; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995) was applied to the issue (Allopenna, Magnuson, & Tanenhaus, 1998).

In the VWP, several objects are shown in a display, and participants are typically instructed to point to or click on one of the objects. As participants listen to the spoken phrase that specifies the target object, their eye movements are recorded. The distribution of fixation proportions over time reveals temporally precise patterns of competition. Specifically, Allopenna et al. (1998) found strong early competition for words sharing onsets (that is, fixations to a beetle when the target was *beaker*) and weaker, later competition for words sharing offsets (with fixation proportion changes lagging about 200 msec behind relevant phonetic information; given the time required to plan and execute an eye movement, this is close to the limit of how short the lag could be). That is, when the target was *beaker*, there was more fixation for *speaker* than for a phonologically unrelated competitor such as *carriage*, but the effect was smaller

and later than for *beetle*. In other words, participants' eye movements were precisely time-locked to the phonological similarity between a presented target word and objects in the display (note that this paradigm is also sensitive to competitors that are not in the display; e.g., Magnuson, Dixon, Tanenhaus, & Aslin, 2007). Related studies have examined in detail the VWP's ability to detect fine-grained effects of phonological similarity. The VWP has been used to track such subtleties as minute differences in coarticulatory cues (Dahan, Magnuson, Tanenhaus, & Hogan, 2001), early and gradual effects of word frequency (Dahan, Magnuson, & Tanenhaus, 2001), and approximately 20-msec differences in vowel duration as a cue to word length (Salverda, Dahan, & McQueen, 2003). Although one must be cautious in avoiding effects of visual similarity (Dahan & Tanenhaus, 2005) and consider carefully how to link the forced-choice nature of the task to theories or models (Tanenhaus, Magnuson, Dahan, & Chambers, 2000), the VWP provides a very sensitive, temporally precise measure of competition during spoken word processing.

If failure to find semantic priming for prime–target pairs with distant or weak semantic relations (Cree et al., 1999; McRae & Boisvert, 1998; Shelton & Martin, 1992) is analogous to failures to find phonological priming for words with matching offsets (Marslen-Wilson & Zwitserlood, 1989), the VWP may provide a more sensitive test that can detect these small effects. Indeed, recently the VWP has been extended to study semantic similarity. Listeners fixate images of semantically (functionally or categorically) related objects more than they do unrelated images (Huettig & Altmann, 2005; Huettig & McQueen, 2007; Huettig, Quinlan, McDonald, & Altmann, 2006; Yee & Sedivy, 2006). So far, these results have been analogous to semantic priming results, although the effects of distantly related items predicted by the Cree et al. attractor network have not been investigated with this paradigm. Thus, applying the VWP to this specific issue has the potential to reveal whether failures to detect priming among distantly related items (Cree et al., 1999; McRae & Boisvert, 1998; Shelton & Martin, 1992) indicate that those attractor model predictions are incorrect, or, instead, that detecting them requires a measure with greater sensitivity. Note that VWP studies of semantic competition go beyond simple analogy to priming, in that these studies can provide new details about the time course of semantic activation and competition. Thus, one might expect that, in the same way that the VWP revealed critical aspects of the time course of phonological competition (Allopenna et al., 1998), it may reveal critical aspects of the time course of semantic competition. In order to generate concrete predictions about the time course of semantic activation and competition, we first conducted simulations with a computational model using distributed feature-based semantic representations.

ATTRACTOR MODEL SIMULATION

Model Architecture and Simulation Design

We used the model developed by Cree and colleagues (Cree et al., 1999; see also O'Connor, Cree, & McRae, 2009), and we would expect similar behavior from other attractor dynamical models of semantic processing (e.g.,
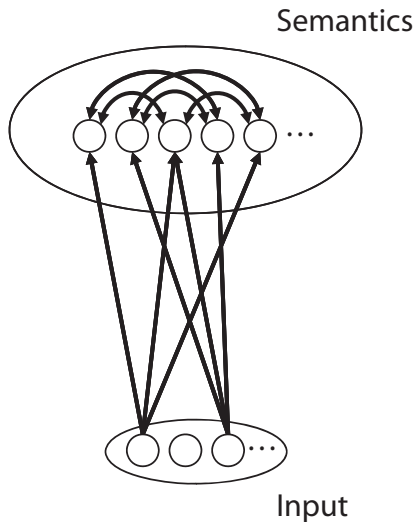
Semantics

Input

Figure 1. Architecture of the nonlinear attractor dynamical model. The word form input layer had 40 units; the semantic feature layer had 2,526 units, corresponding to the 2,526 semantic features in the corpus. Not all connections are shown; where they are, full connectivity was used. Connection weights are established through training.

Plaut & Booth, 2000; Rogers & McClelland, 2004). The model architecture is shown in Figure 1. The "word form" input patterns (analogous to spoken or written word patterns) were created by pseudorandomly selecting 4 out of 40 units to be activated so that no two concepts would have the same input pattern. Target semantic patterns were created by activating the units corresponding to semantic features for each concept from a large corpus of feature norms (McRae et al., 2005). The semantic feature norm corpus contains the responses of 30 participants asked to list up to 10 features for each of 541 concepts. The semantic layer consisted of 2,526 units, each one corresponding to a unique semantic feature in the human subject feature norms. For each concept, units corresponding to features that were produced for that concept by at least 5 participants were activated in the target pattern.

The network was trained to settle to the empirically determined semantic feature pattern for each concept over the course of 20 processing cycles. Following O'Connor et al. (2009), we set learning rate to 0.01 and added momentum (0.9) after the first 10 training epochs. The model was trained using continuous recurrent backpropagation through time (Pearlmutter, 1995) until it correctly activated over 95% of the appropriate semantic feature units (i.e., the model activated over 95% of features that were produced by participants in the feature norming study; by this point, the model also correctly deactivated over 99% of nonproduced features), which was approximately 40 training epochs. Simulations were carried out using MikeNet version 8.02 (www.cnbc.cmu.edu/~mharm/research/tools/mikenet/).

At the end of training, the model was tested on 36 sets of four critical concepts. Each set contained a target concept (the input presented to the model), a near semantic neighbor, a distant semantic neighbor, and an unrelated concept. This design mirrored the design of the priming simulation

and experiment in Cree et al. (1999), which used a target, highly similar prime (near semantic neighbor), less similar prime (distant semantic neighbor), and unrelated concept. Semantic relatedness was based on cosine similarity between feature vectors. Near neighbors had cosine similarity to target greater than .4 and less than 1.0, distant neighbors had cosine similarity to target greater than .1 and less than .4, and nonneighbors had 0 cosine similarity to target (i.e., no common semantic features). These thresholds are comparable to those used in a previous study of near and distant semantic neighbor effects (Mirman & Magnuson, 2008) and were chosen so that distant neighbor similarity would be in the same range as studies that had failed to detect priming effects for distant neighbors and so that there would be enough near neighbors to choose test materials that were matched on other variables.

In anticipation of behavioral experiments, the competitors (near neighbor, distant neighbor, nonneighbor) were matched on familiarity (included in the feature norm corpus; McRae et al., 2005), word frequency (log frequency from the HAL corpus; Lund & Burgess, 1996), and length in phonemes and syllables (the full stimulus list is provided in Appendix A). The top part of Table 1 shows the critical competitor condition means for lexical and semantic variables. To facilitate comparisons with priming studies that failed to find semantic priming for low semantic similarity pairs, we computed cosine similarity for items used in those studies (Cree et al., 1999; McRae & Boisvert, 1998; Shelton & Martin, 1992). Those values are shown in the bottom part of Table 1 and are in the same range as our distant neighbor cosine similarity values.

Results and Discussion

The proximity of the network's semantic state at each processing cycle to the target and competitor concept patterns was evaluated by computing cross-entropy error (CEE) relative to the representation of each concept.[2] Lower CEE indicates greater proximity in semantic space and, hence, greater activation. CEE at each cycle relative to target, near neighbor, distant neighbor, and nonneighbor concepts is plotted in Figure 2 (note that the y-axis has been reversed so that reductions in CEE, which correspond to increases in activation, are higher in the figure). The model exhibited a graded pattern of semantic similarity: CEE was lower to near neighbors than to distant neighbors and was lower to distant neighbors than to nonneighbors. Thus, the model predicts strong activation of near semantic neighbors and weaker activation of distant semantic neighbors.

In addition, there was an interesting difference in the dynamics of activation of near semantic neighbors relative to distant neighbors and nonneighbors. Activation of (proximity to) near neighbors initially gradually increased, peaking at approximately the 11th time cycle ($M = 10.8$, $SD = 5.2$), then decreased. In contrast, activation of (proximity to) distant neighbors and nonneighbors showed a nearly monotonic decrease that was slower for distant neighbors than for nonneighbors (the very small, very early peak is due to the model turning off units that were active as part of the randomly set initialization state).

Table 1
Values for Each Competitor Condition

	Near Neighbors		Distant Neighbors		Nonneighbors	
	M	SD	M	SD	M	SD
Present Experiment						
Cosine similarity	.495	.0915	.218	.0821	.0	.0
Familiarity	4.55	2.08	4.60	2.06	4.56	2.07
Frequency	7.41	1.52	7.48	1.26	8.01	1.55
Number of phonemes	4.58	1.78	4.50	1.38	4.75	1.81
Number of syllables	1.67	0.76	1.67	0.76	1.67	0.76
Previous Priming Studies: Cosine Similarity						
Shelton & Martin (1992) (14 of 36)			.220	.123		
McRae & Boisvert (1998) (27 of 27)	.481	.148	.142	.0601	.0	.0
Cree, McRae, & McNorgan (1999) (11 of 18)	.534	.132	.345	.0529		

Note—Cosine similarity for previous priming studies that failed to find priming for distant neighbors. Values in parentheses indicate how many prime–target pairs were in the feature norm corpus used to compute cosine similarity.

These simulations make two behavioral predictions: The first is that distant semantic neighbors should be more active than nonneighbors (cf. Cree et al., 1999) but not as active as near neighbors. This prediction is not consistent with previous failures to show priming for distantly semantically related concepts, but that could be due to insufficient

sensitivity of the priming paradigm (as in the phonological case; Allopenna et al., 1998). The second prediction is that near neighbors should show a transient peak in activation that is not exhibited by distant neighbors or nonneighbors. We are not aware of any data that speak to this prediction, and testing this prediction requires a behavioral method that provides a temporally precise measure of activation. In the following experiment, we tested these two predictions by examining semantic competition in the VWP.

EXPERIMENT

Method

Participants. Participants were 38 students at the University of Connecticut who reported English as their native language and normal hearing and normal vision (due to technical limitations of the eyetracking equipment, participants wearing glasses or contact lenses were excluded). Participants received course credit.

Materials. Critical stimuli were the same 36 sets of four words used in the simulations; these words differed in concept similarity but were equated on nonsemantic variables (see Table 1). Pairwise t tests showed that none of the control variables differed reliably between conditions, although the nonneighbors had marginally higher word frequency [relative to near neighbors, $t(35) = 2.02$, $p = .051$; relative to distant neighbors, $t(35) = 1.96$, $p = .058$; all other $ts < 1.0$, $ps > .1$]. If this marginal frequency difference had any effect on the fixation data, it would increase fixation of the higher frequency nonneighbors (Dahan, Magnuson, & Tanenhaus, 2001; Magnuson et al., 2007), thus reducing hypothesized semantic similarity effects. All stimuli were produced by a female native speaker of American English in a sound-attenuated room and digitized at 44 kHz. The individual words were edited to eliminate silence at the beginning and end of each sound file.

Procedure. Gaze position and duration were recorded using an ASL 6000 remote eyetracker. Stimulus presentation and response recording were conducted by E-Prime software (Psychological Software Tools, Pittsburgh, PA). Participants were seated with their eyes approximately 27 in. from a 17-in. screen with resolution set to 1,024 × 768 pixels. To ensure that each trial would begin with the participant fixating the neutral central location, participants clicked on a central fixation cross to begin each trial. On each trial, participants saw four images; each image was presented near one of the screen corners, 154 pixels away from the side edges and 115 pixels away from the top and bottom edges (15% of the screen size from the corners). Images had a maximum size of 200 × 200 pixels and were scaled such that at least one dimension was 200 pixels. After a 750-msec preview (to allow for initial fixations that are driven by random factors or visual salience rather than word processing), participants
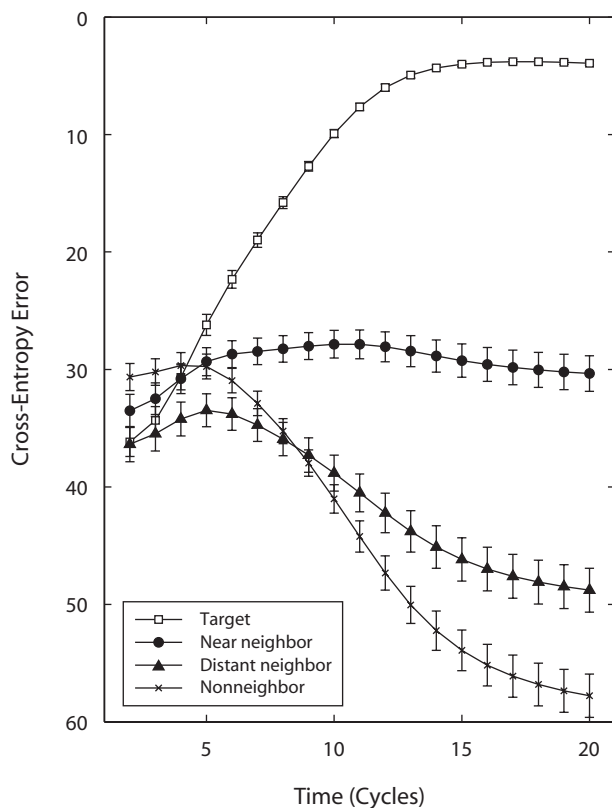

Figure 2. Attractor-based model simulation results. Cross-entropy error at each processing cycle relative to target, near neighbor, distant neighbor, and nonneighbor concepts. Error bars indicate ±1 SE. Lower error indicates greater proximity in semantic space, and thus, greater activation. Note that the y-axis has been reversed so that lower error (greater activation) is higher, to facilitate comparison with the behavioral data.
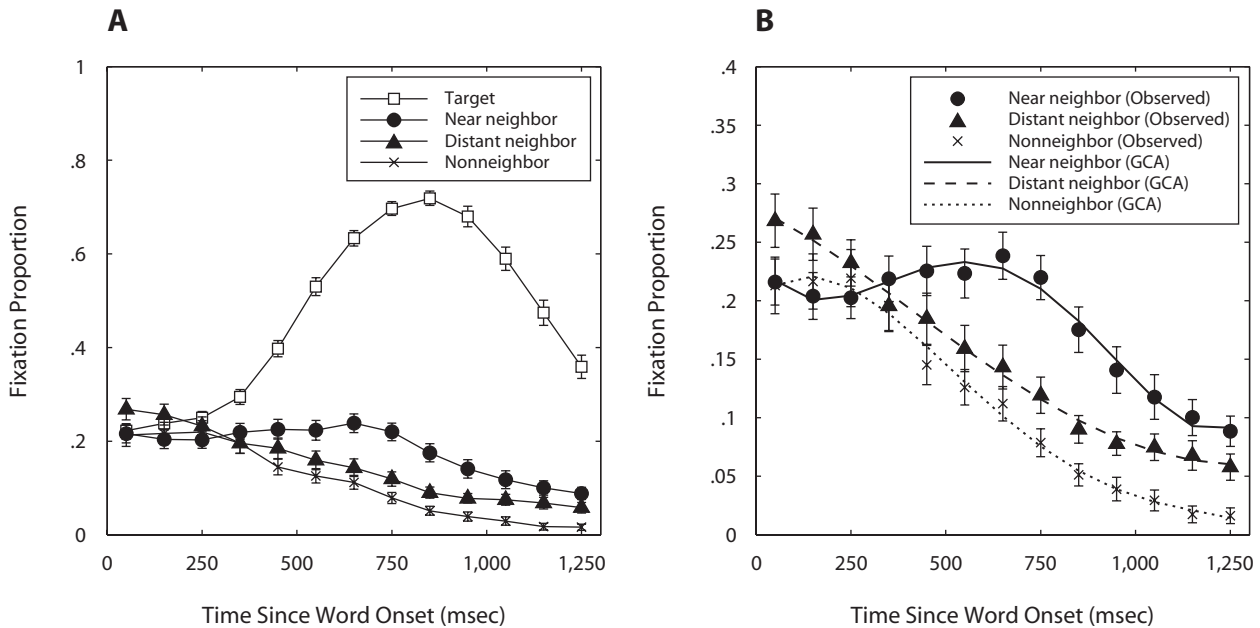
A

B

Figure 3. (A) Fixation proportion for target images (squares), near semantic neighbors (circles), distant semantic neighbors (triangles), and unrelated items (×s). (B) Observed and growth curve analysis (GCA) model fits for competitor fixations. Error bars indicate ±1 *SE*.

heard the target word through headphones and then had to click on the image that corresponded to the target word.

On a critical trial, the target appeared with one of the competitors and two unrelated distractors. There were three counterbalanced lists such that each target occurred once for each participant and in each condition across participants. As a result, each participant had 12 trials in each condition. In addition to the 36 critical trials, there were 48 filler trials on which a target was presented with no related distractors, and the experiment began with 11 practice trials on which feedback was provided.

Results

The left panel of Figure 3 shows average fixation proportions to targets, near neighbors, distant neighbors, and nonneighbors. There was a graded semantic competition effect reflecting the semantic similarity difference between near neighbors, distant neighbors, and nonneighbors. This result converges with previous findings of graded priming effects as a function of degree of featural overlap (Cree et al., 1999) and extends those findings by showing an effect for distant semantic neighbors. The graded pattern of competition demonstrates that the VWP is a sensitive measure of graded semantic similarity effects (see also Huettig & Altmann, 2005; Huettig et al., 2006).

Growth curve analysis (GCA) with orthogonal polynomials was used to quantify differences in the fixation time course for near and distant neighbors relative to nonneighbors (Mirman, Dixon, & Magnuson, 2008). Under the GCA approach to analyzing visual world eyetracking data, there are two (or more) hierarchically related submodels to capture the data pattern. The first submodel, usually called *Level 1*, captures the effect of time on fixation proportions using fourth-order orthogonal polynomials. A fourth-order polynomial is necessary to capture the rise

and fall of fixation probabilities over the course of a trial. Orthogonal polynomials are transformations of natural polynomials that make the individual time terms independent (i.e., remove the correlation between, for example, linear and quadratic time), thus allowing a more precise evaluation of differences in dynamics of processing. Specifically, the intercept term reflects average overall fixation proportion (note that on this approach, the intercept term does not stand for the *y*-intercept, but rather is the average *y*-value of the modeled curve), the linear term reflects a monotonic change in fixation proportion (similar to a linear regression of fixation proportion as a function of time), and the quadratic term reflects an increase followed by a decrease. The cubic and quartic terms tend to capture minor details in the asymptotic tails of the fixation proportion curves and do not have cognitively meaningful interpretations in this context (see Mirman et al., 2008, for details on interpretation of effects on different polynomial terms). Note that effects on the intercept term are equivalent to the standard VWP comparisons of overall fixation proportion; thus, GCA contains both the standard analysis and more sophisticated time course comparisons. The growth curves are shown superimposed on the observed competitor fixation proportions in the right panel of Figure 3, and the full statistical results are in Table 2.

Fixation curves for distant neighbors were different from the nonneighbor fixation curves only in terms of the intercept, reflecting a relatively constant difference across the time course. For near neighbors there were also effects on the linear term (reflecting an overall increase in fixation proportion relative to nonneighbors over the time course) and the quadratic term (reflecting a rise in fixation proportion followed by a decrease in fixation proportion).

Table 2
Results of Growth Curve Analysis (GCA) of Behavioral Data

Term	Near Neighbors			Distant Neighbors			Distant vs. Near Neighbors		
	Estimate	t	$p <$	Estimate	t	$p <$	Estimate	t	$p <$
Intercept	0.070	6.9	.0001	0.360	3.6	.001	−0.034	4.1	.001
Linear	0.124	3.7	.001	0.016	0.5	n.s.	−0.108	3.8	.001
Quadratic	−0.111	6.1	.0001	0.019	1.1	n.s.	0.131	7.1	.0001
Cubic	−0.041	2.3	.05	−0.029	1.6	n.s.	0.013	0.7	n.s.
Quartic	0.066	3.6	.001	0.016	0.9	n.s.	−0.051	2.8	.01

Note—The left and middle sections show results of analyses for near and distant neighbors relative to nonneighbors. The right section shows results of GCA for distant neighbors relative to near neighbors.

Direct comparison of distant and near-neighbor conditions showed significant effects on the intercept, linear, quadratic, and quartic terms, confirming differences in fixation time course for near and distant neighbors.

In addition to the overall differences in fixation proportion, the time course revealed a transient peak in fixation of near neighbors relative to distant neighbors and nonneighbors. This effect is captured by the significant effect of near neighbors on the quadratic term relative to distant neighbors and nonneighbors (see Table 2). That is, according to the GCA, the fixation curves for distant neighbors and nonneighbors were essentially parallel with the distant neighbor fixation curve higher (significant effect only on the intercept term). In contrast, the near-neighbor fixation curve also significantly differed in its curvature (significant effect on the quadratic term as well as the intercept term). The near-neighbor peak occurred relatively early in the time course; namely, the near-neighbor fixation peak occurred before the peak in target fixation [peak timing relative to word onset: $M_{near} = 694.7$ msec, $SE_{near} = 47.8$; $M_{target} = 871.1$ msec, $SE_{target} = 27.5$; $t(37) = 3.32$, $p < .01$].

Discussion

The experiment tested two predictions from simulations of a nonlinear attractor dynamical model of semantic processing. The first prediction was that distantly related semantic neighbors should be partially activated during word comprehension. Previous priming studies had failed to find such an effect (Cree et al., 1999; McRae & Boisvert, 1998; Shelton & Martin, 1992). In contrast, the VWP revealed greater fixations to distant semantic neighbors than to unrelated objects. Consistent with the model prediction, fixation of distant neighbors was greater than unrelated concepts and less than near neighbors. This result is consistent with models that propose graded semantic relations rather than categorical relations. In addition, this result shows that the VWP is more sensitive to semantic effects than semantic priming is, which is an important methodological advance for studying semantic processing.

The second prediction was that near neighbors should show an early, transient peak in fixations. This prediction was also borne out in the behavioral data. Before considering the computational implications of this finding it is important to test whether the behavioral results could be due to task demands or to purely visual similarity rather than conceptual similarity. A number of previous VWP studies have shown that similarity in shape and other visual features can cause competition (Dahan & Tanenhaus, 2005; Huettig & Altmann, 2007; Huettig et al., 2006); thus, it is critical to evaluate to what extent our results are due to visual versus general semantic similarity between targets and competitors.

Task Demands

One possible concern is that the nature of the VWP task might force the final fixation to be on the target object, thereby restricting the competitor peak to be before the target peak. That is, participants may look around until they find the target and then stay fixated on the target until they click on it. Such a task constraint would predict that at the time of a correct response, the fixation should be on the target 100% of the time (or very nearly so, making allowance for measurement error), which would entail that nontarget fixations peak prior to the target peak (since the target would peak at 100% once the participant settled on the target). This was not the pattern in our data: The target object was fixated at the time of the response on 87.6% ($SE = 1.3\%$) of the trials, a value substantially lower than 100%. Further, if this were simply due to measurement error or other factors unrelated to semantic processing, there would be no reason to expect the final fixations to exhibit an effect of semantic relatedness. But there was such an effect: Final fixations were significantly [$t(37) = 2.73$, $p < .01$] more likely to be on a semantically related competitor ($M = 6.35\%$, $SE = 0.87\%$) than on an unrelated competitor ($M = 2.66\%$, $SE = 0.98\%$; the remainder of fixations were to the central fixation cross or other nonobject areas of the display). Recall that these results are only for correct trials, so even when participants click on the correct target, they are not always fixating it at the time of their response, and eye behavior even at the time of the mouse-click decision continues to be influenced by semantic relatedness between the target and the competitor. Thus, the VWP task does not force the final fixation to be on the target, and fixation behavior continues to be affected by semantic similarity all the way to the time of the response.

A weaker form of the task-demands hypothesis is that the nature of the VWP task skews competitor fixations to be early without strictly forcing the final fixation to

be on the target. For example, it may be that that once participants look at the target, they have no need to look at any other images. They may still do so on some small proportion of trials, but this soft constraint makes a later peak highly unlikely (if not strictly impossible). In other words, the decision process involved in driving fixations in the VWP may reduce the likelihood of late competitor fixations, thus creating an early peak in the near-neighbor fixation time course. Formally, this would mean a non-linear relationship between activation and fixation probability such that, as the target becomes highly active, the likelihood of fixating increases much more rapidly than does the likelihood of fixating a less active competitor. The R. D. Luce (1959) choice rule is precisely this type of decision mechanism, and it has been used previously to model decisional aspects of fixation behavior in the VWP. We evaluate whether the Luce choice rule can be solely responsible for producing the early near-neighbor peak in simulations reported below in the Testing Computational Mechanisms section.

Controlling for Visual Similarity

There is no complete and independent way to assess visual similarity, because no established image processing or object recognition machine algorithms exist and human ratings are likely to be influenced by nonvisual semantic similarity (for example, similarity ratings are influenced by category membership; Goldstone, Lippa, & Shiffrin, 2001). To address this issue, we used two complementary measures with different strengths and weaknesses. We based the first measure on the visual features listed in the McRae et al. (2005) feature norms. This measure allowed us to isolate specifically visual similarity from nonvisual semantic similarity, but it is incomplete and does not capture the details of the specific images used in the study. The second measure was based on visual concept–picture similarity ratings. These ratings were based on the specific target concepts and target and nontarget images used in the study, but since human raters are unlikely (or perhaps even unable) to completely ignore the similarity between concepts that is not visual (e.g., that tuba and trombone are both musical instruments), it is likely that these ratings were influenced by nonvisual semantic similarity.

Visual features were identified from among the 2,526 semantic features in the norms on the basis of the brain region taxonomy provided in the norms (see also Cree & McRae, 2003). The color and form/surface categories were chosen because such features would be visible in static images. A small number of features relating to visual variability (e.g., that a balloon can be different col-

ors) and to visual behavior (e.g., that jeans fade) were excluded, because these would not be visible in individual pictures. Visual similarity was then computed as the number and proportion of shared visual features. Both of these measures were considered, because the concepts differed substantially in number of visual features: Proportion of shared visual features abstracted across these differences (e.g., sharing 10 of 20 visual features is equivalent to sharing 5 of 10 visual features) and number of shared visual features captures the fact that some conceptual representations are more reliant on visual features (sharing 10 features is more important than sharing 5 features). Because we did not have a theoretical commitment to one of these measures of visual similarity, we tested both.

In the main experiment, participants performed what is essentially a word–picture matching, so the visual similarity norming study was designed to test the extent to which the specific images used in the experiment were visually similar to the concepts denoted by the target words. That is, for a target word such as *tomato*, we want to know how tomato-like the tomato image was and how tomato-like each competitor image was. This concept–picture visual similarity rating is more relevant in the context of our experiment than is picture–picture similarity, because we need participants to rate the visual similarity of the critical pictures (e.g., tomato, strawberry, potato, magazine) to their mental image induced by the target word (e.g., "tomato"). Visual similarity ratings were collected by presenting an individual image for 200 msec followed by a concept name and asking participants to rate the visual similarity between the image and the concept name on a 5-point scale. Each target concept name was presented with each of the four critical images (target, near neighbor, distant neighbor, unrelated). This approach provides a measure of target-likeness for each critical image. That is, for the target, *tomato*, these ratings provide a measure of similarity to the target word for the critical target image (tomato), near-neighbor image (strawberry), distant neighbor image (potato), and unrelated image (magazine).[3] Twenty-four University of Connecticut undergraduates completed the experiment for course credit. All participants were native English speakers and reported normal or corrected-to-normal vision. The average visual similarity rating for the target pictures was nearly at ceiling (4.6 out of 5; see Table 3), confirming that the target pictures were good representations of target objects.

The three measures of visual similarity (number of shared visual features, proportion of shared visual features, and rated similarity) were very highly correlated (all rs $>$ .5, all ps $<$ .0001). Table 3 shows that, for the

Table 3
Visual Similarity Values

	Target		Near Neighbor		Distant Neighbor		Unrelated	
	M	SD	M	SD	M	SD	M	SD
Number of shared visual features	5.4	2.4	2.6	1.2	1.2	1.1	0.0	0.0
Proportion of shared visual features	1.0	–	0.50	0.26	0.21	0.22	0.0	0.0
Rated similarity (1–5)	4.6	0.36	2.3	0.89	1.2	0.19	1.1	0.08

Table 4
Effect of Adding Groups of Terms on Model Fit

	−2LL	ΔLL	p <	+Semantic Similarity −2LL	+Semantic Similarity ΔLL	+Semantic Similarity p <	−Visual Similarity ΔLL	−Visual Similarity p <
Base model	−4,172	–	–	−4,553	381	.0001	–	–
Number of shared visual features	−4,344	172	.0001	−4,565	221	.0001	11.34	.05
Proportion of shared visual features	−4,319	147	.0001	−4,560	241	.0001	6.13	n.s.
Visual similarity rating	−4,353	181	.0001	−4,562	209	.0001	8.44	n.s.

Note—The right section shows results for the base model and improvements in model fit when individual visual similarity terms are added. The middle section shows improvements in model fit when semantic similarity condition terms are added. The left section shows the effect of removing visual similarity terms.

competitors, the number and proportion of visual features shared with the target and the rated visual similarity to the target followed the overall semantic similarity pattern. This is not surprising; the critical question is whether this pattern can account for the semantic similarity effect. To test that, we conducted a series of by-items GCAs of the fixation proportion data (analogous to the by-subjects GCA reported above). In order to evaluate the significance of visual and semantic similarity, we examined the change in the deviance statistic −2LL (minus 2 times the log-likelihood). The *change in deviance*, ΔLL, is distributed as chi-square, with degrees of freedom equal to the number of parameters added to the model. The starting point was a base model that included just the time terms and the item effects on time terms; that is, the base model contained neither visual similarity nor semantic similarity effects. When visual similarity terms were added, they provided a significant improvement in model fit (ΔLL > 140, p < .0001, for each measure of visual similarity). When semantic similarity condition terms were added to the model already containing visual similarity terms, they provided significant additional improvement in model fit (ΔLL > 200, p < .0001, for each measure of visual similarity). That is, semantic similarity condition accounted for significant additional variance beyond the variance captured by visual similarity. In fact, when visual similarity terms were removed from this full model, the decrease in model fit was small and reliable only for the number of shared visual features measure (ΔLL = 11.34, p < .05). The goodness of fit results, summarized in Table 4, indicate that, for competitor fixation time course in the present experiment, visual similarity to the target represented only a subset of semantic similarity. See Appendix B for detailed results of by-items growth curve analyses after visual similarity measures were included.

These analyses demonstrate that fixations to the semantic competitors were driven primarily by semantic, not visual, similarity. The behavioral results were consistent with both of the critical predictions of the nonlinear attractor dynamical model: a graded pattern of semantic competition based on semantic relatedness and an early, transient peak for the near semantic neighbor. In order to examine what these patterns indicate about the computational mechanisms of semantic processing, we explored what is required for different computational models to account for these results.

TESTING COMPUTATIONAL MECHANISMS

The behavioral data closely matched the predictions from the attractor dynamical model, suggesting that this model correctly captures important aspects of the computation of word meaning; however, it is not clear what details of the computational model give rise to the observed patterns. One way to pinpoint the critical computational aspect is to construct alternative models that also fit the behavioral data and examine their computational similarities. To that end, we considered two alternative computational frameworks: a localist spreading activation model and a simple decision model.

Localist Spreading Activation Model

In spreading activation models, when a concept node becomes active, activation spreads to all of its related concepts (Collins & Loftus, 1975; McNamara, 1992). If the link strength is proportional to relatedness between the concepts, these models straightforwardly predict the graded semantic similarity effects found in the present eyetracking study and in priming studies.

In standard spreading activation models (Figure 4, left panel) external input gradually activates a target concept and this activation spreads to connected (related) concepts. Unchecked, this would lead to full activation of all units in the network, so one more assumption is required: that unit activations decay when external input is removed. With decay included, the target concept node will become active, then the activation will decay, producing a transient peak. Because the target concept node drives the activation of related nodes, the near-neighbor activation profile will necessarily lag behind it: The neighbor's activation peak will always occur after the target's activation peak. The situation is analogous to a sequence of falling dominoes: The first domino (target) causes the second one (neighbor) to fall, but the first necessarily falls first and reaches its maximum velocity first, and so on. Simulations of a simple spreading activation network consisting of a target unit (that received the external input excitation) and a neighbor unit (that was activated by spreading activation from the target unit) confirmed that this pattern emerges independent of parameter settings (e.g., connection weights ranging from .05 to 1.0 were tested). Note that adding more concept units may change the magnitude of the activations, but not the temporal pattern of activations:
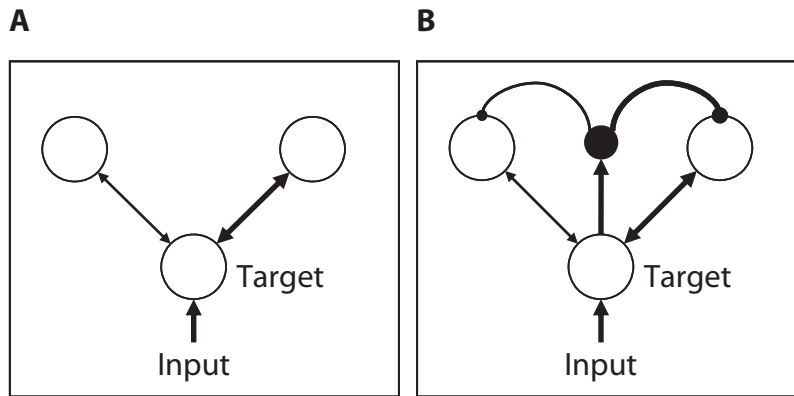
A **B**

Figure 4. Two types of spreading activation models. (A) Standard spreading activation model with localist representations and only excitatory connections. Connection strength is indicated by line thickness. The target receives external input and spreads activation to related (connected) nodes. (B) A localist spreading activation model with an inhibitory node that is activated by the target and inhibits the other nodes in the model (similar inhibitory nodes for each concept node in the model are not shown).

Neighbor activations must lag behind target activations. Thus, standard spreading activation networks predict that neighbor activation will peak after target activation has peaked, a pattern that is not consistent with the results of our eyetracking experiment.

A spreading activation model can exhibit an early, transient neighbor activation peak if the standard architecture is extended to include inhibitory nodes that are activated by individual concept nodes and that inhibit all other concept nodes (Figure 4, right panel). Note that simply adding lateral inhibitory connections to a standard spreading activation network will merely decrease the net connection strength between units without changing the qualitative dynamics of the units. For the extended version of the spreading activation model, external input will activate the target concept node, neighbor nodes will be initially activated by the excitatory connections from the target, but as the inhibitory unit becomes more active, this unit will inhibit the neighbors. When properly parameterized, these networks exhibit stable activation of the target concept and early activation peaks for neighbor nodes. This kind of separation of excitatory and inhibitory pathways can be found in biological neural systems (e.g., inhibitory interneurons; see Markram et al., 2004, for a review). These simulations suggest that a combination of excitatory and inhibitory interactions between units is necessary for this model to produce the early neighbor activation peak.

Decision Model

One advantage of the VWP is that it minimizes the kind of metalinguistic task demands that arise in lexical decision or semantic categorization tasks. Nonetheless, the behavioral data necessarily reflect some amount of decisional processing, at the very least at the level of oculomotor control. As discussed above, it is possible that decision processes produce the early near-neighbor fixation peak. Formulating a detailed decision model for this paradigm is outside the scope of this report; however, a few basic prin-

ciples are enough to formulate a simple decision model. First, the decision likelihood should be monotonically related to semantic activation such that images corresponding to semantic representations that are more active should be more likely to be fixated. This principle reflects the excitatory aspect of decision processes: Increased activation increases fixation likelihood for the corresponding image. Second, when a single representation has reached a sufficiently high level of activation (i.e., the target has been recognized), fixations of competitor images should be very low. This second principle reflects the inhibitory aspect of decision processes: Greater activation of one representation decreases fixation likelihood for competitor images. These basic principles are encapsulated by the R. D. Luce (1959) choice rule, which has been widely used to model decision-level aspects of spoken word processing (e.g., McClelland & Elman, 1986), particularly for the VWP (e.g., Allopenna et al., 1998; Dahan, Magnuson, & Tanenhaus, 2001; Dahan, Magnuson, Tanenhaus, & Hogan, 2001).

According to the R. D. Luce (1959) choice rule, the probability of response i from a set of alternatives indexed by j is

$$p(\mathrm{R}_i) = \frac{e^{ka_i}}{\sum_j e^{ka_j}},$$

where a_i is the activation level of representation i, and k is a constant. Under the Luce rule, representations with higher activation are more likely to produce the response (i.e., to be fixated) and all response probabilities add to 1.0 (i.e., the set of alternatives accounts for all fixations). The exponential factor in the Luce rule causes a nonlinear relationship between relative activation and relative likelihood of response (fixation); thus, when a single representation is relatively highly active, other representations have a very low likelihood of being fixated.

The strongest test of whether this decision mechanism can produce an early neighbor fixation peak is for the

input to the decision mechanism not to have any activation peaks. This way, any peaks in the decision model's output are necessarily due to the decision model. To that end, we examined a simple case in which activation for a target representation increases linearly, activation for a semantic neighbor increases at a fixed ratio of target activation, and activation of two unrelated competitors is fixed at a low, nonzero value (Figure 5B, top left panel). This formulation captures the four-alternative forced choice aspect of the VWP (participants must fixate one of four objects in the display) and does not include an underlying activation peak for the neighbor; thus, a peak in response probability would be due to the R. D. Luce (1959) choice model. Since the input target activation is increasing linearly, target fixation probability increases until it asymptotes at 1.0 and the neighbor peak can only occur early with respect to a target peak. That is, for the decision model, the critical question is whether it produces a neighbor peak or not; the timing of the peak is constrained by the simplified input to the model. This formulation of the decision model has two free parameters: the target–neighbor activation ratio and the constant, k. A full exploration of this two-dimensional parameter space revealed that the Luce choice rule produces three distinct patterns of competitor response probability over time:

1. A gradual, monotonic increase in response probability for both the target and the neighbor (Figure 5B, top right panel). This pattern reflects a case of high ambiguity (high activation of the neighbor relative to the target) that

persists throughout the time course due to low selectivity (low k). If the target and neighbor activation are allowed to continue to rise, the difference between them eventually becomes large enough that the target draws most of the responses, and the neighbor response probability begins to decrease, producing a transient peak. However, this peak falls outside the scope of the activation values tested.

2. An initial increase in response probability followed by a decrease—that is, a transient peak in response probability (Figure 5B, bottom left panel). This pattern emerges when the activation of the neighbor is moderately high relative to the target, so the probability of fixations initially increases, but as the target becomes more active, it draws an increasingly higher proportion of fixations, and the fixations of the neighbor drop, producing the transient peak.

3. A gradual, monotonic decrease in response probability (Figure 5B, bottom right panel). This pattern emerges when the activation of the neighbor is low relative to the activation of the target, so it does not exhibit the initial increase in fixation probability.

Figure 5A shows the parameter space with curves that represent the dividing lines between the patterns. When correctly parameterized, the R. D. Luce (1959) choice model predicts the observed transient peak in fixation likelihood for near neighbors.

Discussion

We examined three distinct computational frameworks to determine whether they predict an early, transient peak
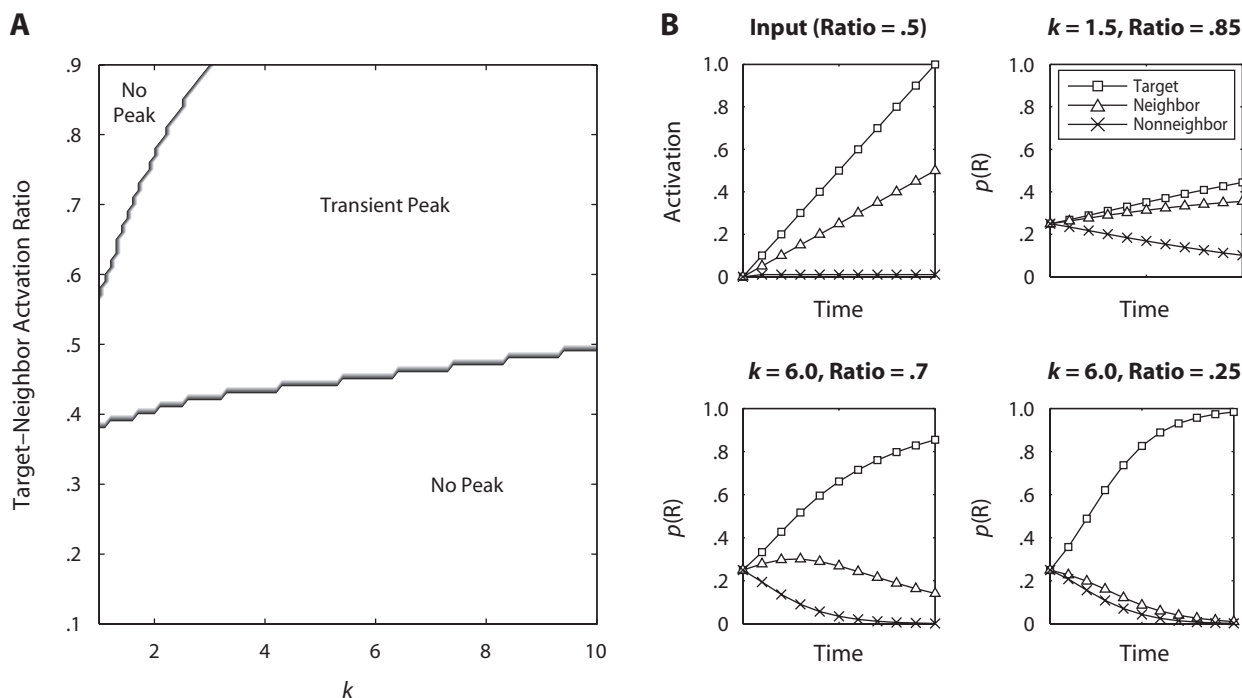

Figure 5. (A) R. D. Luce (1959) choice model parameter space. The curves represent the dividing lines between the three observed patterns. (B) Top left panel: Example activation patterns that served as input to the decision model. Top right panel: Example point from the parameter space, demonstrating monotonic increase in neighbor response proportion (no peak). Bottom left panel: Example point from the parameter space, demonstrating early, transient peak in neighbor response proportion. Bottom right panel: Example point from the parameter space, demonstrating monotonic decrease in neighbor response proportion (no peak).

in near-neighbor fixation. A nonlinear attractor dynamical model naturally produced such a peak. A localist spreading activation model produced such a peak when the model included separate excitatory and inhibitory pathways. A decision model also produced a peak, but in a restricted parameter regime. The common pattern across these three computational mechanisms is the combination of excitatory and inhibitory mechanisms. This point is most clear when the spreading activation model is considered: With only excitatory connections, the model could not account for the early peak; but with the addition of inhibitory connections, the early peak emerged. In the case of the decision model, the excitatory aspect was built into the input (target and neighbor activation increased linearly) and the decision rule acted as an inhibitory mechanism.

In the case of the nonlinear attractor dynamical model, both the excitatory and inhibitory effects were emergent properties of partially overlapping distributed representations. For example, in the framework of feature-based semantic representations, *lion* and *tiger* concepts are similar because they share many features (e.g., *is a predator*, *has teeth*, *lives in wilderness*, *is used in circuses*) but not all features (e.g., a lion has a mane; a tiger has stripes). The shared features give rise to excitation: Activation of the semantic representation of *lion* partially activates *tiger*, because some features of *tiger* (the ones it shares with *lion*) are activated. The features that are not shared give rise to inhibition: Activation of *has mane* is inconsistent with tiger, as is nonactivation of *has stripes*. In general, for attractor dynamical models of semantics, the appropriate patterns of excitatory and inhibitory connections emerge as a necessary consequence of the training regime (see Rogers & McClelland, 2004, particularly chap. 4, for discussion of development of concept attractors in such networks).

In sum, examination of three computational frameworks showed that a combination of excitatory and inhibitory mechanisms is necessary to produce the early peak in neighbor activation observed in the fixation time course data. This combination can be an emergent property (as in attractor dynamical models based on distributed representations) or explicit (as in localist spreading activation models), or it can be due to competition in the decision mechanism or possibly to other mechanisms. Whatever general computational framework one chooses, the present behavioral data impose the constraint that excitatory and inhibitory mechanisms must be present in the formulation of any model.

CONCLUSIONS

Semantic similarity is a critical component of understanding the organization of semantic knowledge. The present experiment used the VWP to investigate the time course of semantic similarity effects as a step toward uncovering the dynamics of semantic processing during spoken word recognition. The likelihood of fixation was predicted by degree of semantic similarity to the target concept. Post hoc analyses of recent studies that used the VWP to examine semantic similarity effects have suggested the presence of such graded effects (e.g., Huettig & Altmann, 2005), but this is the first study in which multiple levels of semantic similarity were tested explicitly. Critically, distant semantic neighbors with similarity too low to produce semantic priming (based on previous studies of semantic priming under low similarity conditions) elicited robust semantic competition, as revealed by the distribution of fixation proportions over time. That is, listeners were more likely to look at images of distant semantic neighbors than of unrelated concepts, consistent with the gradient semantic activation predicted by the attractor network as a function of similarity distance. This finding shows that the VWP provides an important tool for investigating semantic structure, because it can reveal effects that may not be detected in priming experiments.

The time course revealed a novel behavioral finding: a relatively early peak in fixation of near semantic neighbors that did not occur for distant neighbors or nonneighbors. Simulations of an attractor dynamical model of word comprehension showed that this pattern is a natural consequence of attractor dynamics. An early peak in neighbor activation can also emerge in a spreading activation model, if the standard architecture is extended to include both excitatory and inhibitory pathways, and in a properly parameterized decision model. The common computational aspect of all three models that produce the early peak is that they have a combination of excitatory and inhibitory processes. In the attractor dynamical model, excitation and inhibition are emergent properties of distributed semantic representations, the spreading activation requires explicit excitatory and inhibitory pathways, and the decision model assumes excitatory dynamics at the input level and provides a postperceptual inhibitory mechanism that can produce the observed peak pattern.

In summary, our behavioral results supported predictions of gradient semantic activation and competition down to even very low levels of featural overlap that previously had failed to result in detectable semantic priming. Methodologically, this result indicates that the VWP is a powerful tool that can detect semantic similarity effects that are too small to be detected with semantic priming, the technique that has dominated in this field of study for 30 years. For theories of semantic memory, this result indicates that the similarity between concepts is graded rather than categorical, which is consistent with distributed models of semantic cognition rather than with localist models. In addition, our behavioral results introduced new constraints regarding the time course of activation of target concepts and of semantically related concepts. Our computational explorations showed that, although very different implementations are viable, a combination of excitatory and inhibitory dynamics is necessary to generate the appropriate time course. Thus, the combination of computational and behavioral investigations led to novel

insights into the dynamics of semantic activation and competition.

REFERENCES

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory & Language*, **38**, 419-439.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral & Brain Sciences*, **22**, 577-660.

Chiarello, C., Burgess, C., Richards, L., & Pollock, A. (1990). Semantic and associative priming in the cerebral hemispheres: Some words do, some words don't . . . sometimes, some places. *Brain & Language*, **38**, 75-104.

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, **82**, 407-428.

Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language: A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology*, **6**, 84-107.

Cree, G. S., & McRae, K. (2003). Analyzing the factors underlying the structure and computation of the meaning of chipmunk, cherry, chisel, cheese, and cello (and many other such concrete nouns). *Journal of Experimental Psychology: General*, **132**, 163-201.

Cree, G. S., McRae, K., & McNorgan, C. (1999). An attractor model of lexical conceptual processing: Simulating semantic priming. *Cognitive Science*, **23**, 371-414.

Crystal, T. H., & House, A. S. (1990). Articulation rate and the duration of syllables and stress groups in connected speech. *Journal of the Acoustical Society of America*, **88**, 101-112.

Dahan, D., Magnuson, J. S., & Tanenhaus, M. K. (2001). Time course of frequency effects in spoken-word recognition: Evidence from eye movements. *Cognitive Psychology*, **42**, 317-367.

Dahan, D., Magnuson, J. S., Tanenhaus, M. K., & Hogan, E. M. (2001). Subcategorical mismatches and the time course of lexical access: Evidence for lexical competition. *Language & Cognitive Processes*, **16**, 507-534.

Dahan, D., & Tanenhaus, M. K. (2005). Looking at the rope when looking for the snake: Conceptually mediated eye movements during spoken-word recognition. *Psychonomic Bulletin & Review*, **12**, 453-459.

Goldstone, R. L., Lippa, Y., & Shiffrin, R. M. (2001). Altering object representations through category learning. *Cognition*, **78**, 27-43.

Hines, D., Czerwinski, M., Sawyer, P. K., & Dwyer, M. (1986). Automatic semantic priming: Effect of category exemplar level and word association level. *Journal of Experimental Psychology: Human Perception & Performance*, **12**, 370-379.

Huettig, F., & Altmann, G. T. M. (2005). Word meaning and the control of eye fixation: Semantic competitor effects and the visual world paradigm. *Cognition*, **96**, B23-B32.

Huettig, F., & Altmann, G. T. M. (2007). Visual-shape competition during language-mediated attention is based on lexical input and not modulated by contextual appropriateness. *Visual Cognition*, **15**, 985-1018.

Huettig, F., & McQueen, J. M. (2007). The tug of war between phonological, semantic and shape information in language-mediated visual search. *Journal of Memory & Language*, **57**, 460-482.

Huettig, F., Quinlan, P. T., McDonald, S. A., & Altmann, G. T. M. (2006). Models of high-dimensional semantic space predict language-mediated eye movements in the visual world. *Acta Psychologica*, **121**, 65-80.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, **104**, 211-240.

Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear & Hearing*, **19**, 1-36.

Luce, R. D. (1959). *Individual choice behavior: A theoretical analysis*. New York: Wiley.

Lund, K., & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instruments, & Computers*, **28**, 203-208.

Magnuson, J. S., Dixon, J. A., Tanenhaus, M. K., & Aslin, R. N. (2007). The dynamics of lexical competition during spoken word recognition. *Cognitive Science*, **31**, 133-156.

Markram, H., Toledo-Rodriguez, M., Wang, Y., Gupta, A., Silberberg, G., & Wu, C. (2004). Interneurons of the neocortical inhibitory system. *Nature Reviews Neuroscience*, **5**, 793-807.

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition*, **25**, 71-102.

Marslen-Wilson, W. D., & Zwitserlood, P. (1989). Accessing spoken words: The importance of word onsets. *Journal of Experimental Psychology: Human Perception & Performance*, **15**, 576-585.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, **18**, 1-86.

McNamara, T. P. (1992). Priming and constraints it places on theories of memory and retrieval. *Psychological Review*, **99**, 650-662.

McRae, K., & Boisvert, S. (1998). Automatic semantic similarity priming. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **24**, 558-572.

McRae, K., Cree, G. S., Seidenberg, M. S., & McNorgan, C. (2005). Semantic feature production norms for a large set of living and nonliving things. *Behavior Research Methods*, **37**, 547-559.

Mirman, D., Dixon, J. A., & Magnuson, J. S. (2008). Statistical and computational models of the visual world paradigm: Growth curves and individual differences. *Journal of Memory & Language*, **59**, 475-494.

Mirman, D., & Magnuson, J. S. (2008). Attractor dynamics and semantic neighborhood density: Processing is slowed by near neighbors and speeded by distant neighbors. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **34**, 65-79.

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (2004). The University of South Florida free association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments, & Computers*, **36**, 402-407.

O'Connor, C. M., Cree, G. S., & McRae, K. (2009). Conceptual hierarchies in a flat attractor network: Dynamics of learning and computations. *Cognitive Science*, **33**, 665-708. doi:10.1111/j.1551 -6709.2009.01024.x

Pearlmutter, B. A. (1995). Gradient calculation for dynamic recurrent neural networks: A survey. *IEEE Transactions on Neural Networks*, **6**, 1212-1228.

Plaut, D. C., & Booth, J. R. (2000). Individual and developmental differences in semantic priming: Empirical and computational support for a single-mechanism account of lexical processing. *Psychological Review*, **107**, 786-823.

Reppen, R., Ide, N., & Suderman, K. (2005). *American National Corpus*. Philadelphia: Linguistic Data Consortium. http:// americannationalcorpus.org/.

Rogers, T. T., & McClelland, J. L. (2004). *Semantic cognition: A parallel distributed processing approach*. Cambridge, MA: MIT Press.

Salverda, A. P., Dahan, D., & McQueen, J. M. (2003). The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. *Cognition*, **90**, 51-89.

Shelton, J. R., & Martin, R. C. (1992). How semantic is automatic semantic priming? *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **18**, 1191-1210.

Steyvers, M., & Tenenbaum, J. B. (2005). The large-scale structure of semantic networks: Statistical analyses and a model of semantic growth. *Cognitive Science*, **29**, 41-78.

Tanenhaus, M. K., Magnuson, J. S., Dahan, D., & Chambers, C.

(2000). Eye movements and lexical access in spoken-language comprehension: Evaluating a linking hypothesis between fixations and linguistic processing. *Journal of Psycholinguistic Research*, **29**, 557-580.

TANENHAUS, M. K., SPIVEY-KNOWLTON, M. J., EBERHARD, K. M., & SEDIVY, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, **268**, 1632-1634.

VIGLIOCCO, G., VINSON, D. P., LEWIS, W., & GARRETT, M. F. (2004). Representing the meanings of object and action words: The featural and unitary semantic space hypothesis. *Cognitive Psychology*, **48**, 422-488.

YEE, E., & SEDIVY, J. C. (2006). Eye movements to pictures reveal transient semantic activation during spoken word recognition. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **32**, 1-14.

NOTES

1. In English, the average syllable duration is 200–250 msec (Crystal & House, 1990), and the frequency-weighted average word length is 2.3 syllables (American National Corpus; Reppen, Ide, & Suderman, 2005), so the average word duration is approximately 0.5 sec, producing a speech rate of approximately 120 words/min.

2. CEE was computed using a reachable threshold of .1, which means that error was computed relative to a target value of .1 for units that were supposed to be off and a target value of .9 for units that were supposed to be on. Units within .1 of the actual target value were considered as having the correct activation.

3. One might worry that this task may underestimate accidental or nonsemantic surface similarities between our target and competitor pictures. Concept–picture similarity ratings provide an index of how much a physical image of one concept resembles a mental image of another, which provides the closest match to the task demands of the experiment. A picture–picture visual similarity task would provide more information about the physical similarity of the images, but fixation behavior in the VWP is thought to be driven primarily by similarity between images in the display and the active mental representations, not similarity among images in the display. That is, one predicts fixations of a cat because the target word is *dog*, not because there is a dog in the display. Finally, although different visual similarity norming tasks are likely to produce somewhat different results, they are also likely to be very highly correlated (e.g., the high correlation between shared visual features and concept–picture similarity norming), so it is unlikely that a different norming task would produce a substantially different control variable for this analysis.

APPENDIX A
Critical Items

Target	Near Neighbor	Distant Neighbor	Nonneighbor
bayonet	machete	tomahawk	colander
beans	peas	pear	jar
blueberry	raspberry	pineapple	microscope
broccoli	celery	banana	envelope
buffalo	caribou	elephant	clarinet
bus	van	bike	ball
cake	pie	pear	stone
cat	dog	stool (furniture)	doll
cheetah	zebra	bison	tripod
crab	clam	cod	clamp
crow	goose	wasp	snail
dagger	hatchet	shotgun	cello
deer	fox	bear	skis
dove	swan	bat (animal)	hoe
eagle	blackbird	tiger	trumpet
elk	hare	chimp	urn
falcon	partridge	ostrich	sardine
grape	peach	pear	rope
lion	tiger	beaver	crowbar
moose	hare	chimp	sword
owl	hawk	swan	crane (machine)
panther	leopard	zebra	baton
pants	shirt	socks	fork
peacock	blackbird	giraffe	rifle
pelican	flamingo	crocodile	revolver
penguin	starling	otter	missile
pheasant	starling	catfish	cannon
pistol	shotgun	crossbow	banjo
sheep	goat	rat	yacht
sparrow	blackbird	beetle	anchor
spear	bow (weapon)	shield	crown
squid	eel	seal	tank (army)
stork	finch	hare	wand
tomato	strawberry	potato	magazine
trombone	tuba	bagpipe	scooter
truck	van	couch	cheese

Note—In the experiment, only a single critical competitor was presented along with the target on each trial. The choice of competitor presented for each target was counterbalanced across participants.

APPENDIX B

Tables B1–B4 show the parameter estimates and significance values for semantic similarity condition effects on the orthogonal time terms in the different models. The by-items analysis results (Table B1) mirrored the results of the by-subjects analysis (Table 2, main text); specifically, there were significant effects of neighbor condition on the intercept and quadratic terms. Tables B2–B4 show that, even after visual similarity measures were included, semantic neighbor condition still had significant effects on growth curve model terms, particularly the intercept and quadratic terms. These results show that nonsemantic visual similarity cannot account for the competition effects found in the experiment.

Table B1
Results of By-Items Growth Curve Analysis

Term	Near Neighbors			Distant Neighbors			Distant vs. Near Neighbors		
	Estimate	t	$p <$	Estimate	t	$p <$	Estimate	t	$p <$
Intercept	0.072	13.0	.0001	0.034	6.0	.0001	−0.037	7.3	.0001
Linear	0.128	5.8	.0001	0.019	0.9	n.s.	−0.110	5.5	.0001
Quadratic	−0.099	8.9	.0001	0.037	3.4	.001	0.137	12.4	.0001
Cubic	−0.032	2.9	.01	−0.028	2.5	.05	0.004	0.3	n.s.
Quartic	0.043	3.9	.001	−0.003	0.3	n.s.	−0.046	4.2	.0001

Table B2
**Effect of Semantic Competitor Condition After Number
of Shared Visual Features Was Included in the Model**

Term	Near Neighbors			Distant Neighbors			Distant vs. Near Neighbors		
	Estimate	t	$p <$	Estimate	t	$p <$	Estimate	t	$p <$
Intercept	0.057	5.5	.0001	0.028	4.0	.001	−0.040	5.1	.0001
Linear	0.174	4.3	.0001	0.040	1.5	n.s.	−0.110	3.6	.001
Quadratic	−0.090	4.4	.0001	0.042	3.0	.01	0.123	7.3	.0001
Cubic	−0.069	3.4	.001	−0.045	3.3	.001	0.044	2.6	.01
Quartic	0.067	3.3	.01	0.008	0.6	n.s.	−0.055	3.2	.01

Table B3
**Effect of Semantic Competitor Condition After Proportion
of Shared Visual Features Was Included in the Model**

Term	Near Neighbors			Distant Neighbors			Distant vs. Near Neighbors		
	Estimate	t	$p <$	Estimate	t	$p <$	Estimate	t	$p <$
Intercept	0.077	8.0	.0001	0.037	5.6	.0001	−0.050	7.4	.0001
Linear	0.085	2.3	.05	0.0005	0.02	n.s.	−0.071	2.7	.01
Quadratic	−0.121	6.4	.0001	0.028	2.2	.05	0.162	10.4	.0001
Cubic	−0.042	2.2	.05	−0.033	2.5	.05	0.005	0.3	n.s.
Quartic	0.057	3.0	.01	0.003	0.2	n.s.	−0.061	4.0	.0001

Table B4
**Effect of Semantic Competitor Condition After Visual Similarity Rating
Was Included in the Model**

Term	Near Neighbors			Distant Neighbors			Distant vs. Near Neighbors		
	Estimate	t	$p <$	Estimate	t	$p <$	Estimate	t	$p <$
Intercept	0.067	7.4	.0001	0.034	5.9	.0001	−0.038	4.6	.0001
Linear	0.091	2.6	.01	0.014	0.6	n.s.	−0.107	3.3	.001
Quadratic	−0.114	6.4	.0001	0.035	3.1	.01	0.172	9.5	.0001
Cubic	−0.026	1.4	n.s.	−0.027	2.4	.05	0.007	0.4	n.s.
Quartic	0.061	3.4	.001	−0.0007	0.1	n.s.	−0.079	4.4	.0001