COGNITIVE SCIENCE Vol 23 (3) 1999, pp. 371–414 ISSN 0364-0213

An Attractor Model of Lexical Conceptual Processing: Simulating Semantic Priming

GEORGE S. CREE, KEN MCRAE, AND CHRIS MCNORGAN

University of Western Ontario

An attractor network was trained to compute from word form to semantic representations that were based on subject-generated features. The model was driven largely by higher-order semantic structure. The network simulated two recent experiments that employed items included in its training set (McRae and Boisvert, 1998). In Simulation 1, short stimulus onset asynchrony priming was demonstrated for semantically similar items. Simulation 2 reproduced subtle effects obtained by varying degree of similarity. Two predictions from the model were then tested on human subjects. In Simulation 3 and Experiment 1, the items from Simulation 1 were reversed, and both the network and subjects showed minimally different priming effects in the two directions. In Experiment 2, consistent with attractor networks but contrary to a key aspect of hierarchical spreading activation accounts priming was determined by featural similarity rather than shared superordinate category. It is concluded that semantic-similarity priming is due to featural overlap that is a natural consequence of distributed representations of word meaning.

I. INTRODUCTION

Understanding the computation of word meaning is central to the development of a complete account of language comprehension. Accordingly, a great deal of research investigating the structure of semantic memory and the mechanisms governing the performance on semantic tasks has been conducted. During the last 30 years, the study of semantic memory has been dominated by both network theories, such as those of Collins and Loftus (1975), and feature theories, such as those of Smith, Shoben, and Rips (1974). Recently, a number of researchers have extended feature-based theories by instantiating them in distributed connectionist attractor networks (e.g., Becker, Moscovitch, Behrmann,

Direct all correspondence to: Ken McRae, Department of Psychology, Social Science Centre, University of Western Ontario, London, Ontario, N6A 5C2; E-Mail: mcrae@uwo.ca

& Joordens, 1997; Hinton & Shallice, 1991; Masson, 1995; McRae, de Sa, & Seidenberg, 1997). This research has led to insightful accounts of a number of semantic memory phenomena in both normal and neurologically impaired individuals. In this article, we furthered these investigations by presenting an attractor model of the computation of word meaning. In particular, we focused on semantic priming between words that have similar meanings, but are not normatively associated, such as *truck* and *van* or *eagle* and *hawk*. The semantic representations used in the model were taken from the feature production norms of McRae et al. (1997). The use of norms generated by subjects is an important step in modeling because recent experiments have shown that the degree of priming depends on the degree of featural overlap (McRae & Boisvert, 1998; hereafter referred to as MB). Thus, given the constraints imposed by the empirically derived representations, the model must show appropriately sized effects. This contrasts with previous models in this domain in which representational similarity was a free parameter.

In this article, we present two simulations of recent studies of semantic-similarity priming. Two human behavioral experiments that test predictions generated from the model are then presented. The first tests the prediction that the effects of semantic-similarity priming should be basically equivalent in both prime-target directions. The second experiment extends the work of MB by contrasting accounts of semantic-similarity priming that are based on featural similarity or shared superordinate category.

Attractor Networks of Lexical Computations

Connectionist attractor networks are a class of parallel distributed processing models that are being used with increasing frequency in cognitive science. These networks are interactive, cyclical models that represent patterns in a distributed fashion. In an attractor network, a learned pattern is represented as a stable state in a multidimensional space (for a detailed discussion of attractor networks of lexical processing, see Plaut & Shallice, 1993). For example, when computing the meaning and phonology of a word given its orthographic form, units interact and update their states repeatedly until the pattern of activation across all units ceases to change. At this point, the network has reached a stable state, and the corresponding pattern of activation is called an attractor. The set of representational states surrounding each attractor is known as its basin of attraction. Basins of attraction play a key role in the computational dynamics of these models because, if the network is within a learned pattern's basin, it will settle to the corresponding attractor state if no further input intervenes and if processing is deterministic. Consistent with these processing dynamics, learning involves carving the multidimensional state space into attractor basins so that the network settles into the correct stable state for each input.

Attractor networks have been used to account for several behavioral phenomena that are associated with normal and impaired individuals' performance of lexical tasks. Much of this work has dealt with simulating how people pronounce printed words aloud. For example, Plaut, McClelland, Seidenberg, and Patterson (1996) implemented an attractor model of the computation of phonology from orthography, which extended the seminal work of Seidenberg and McClelland (1989) who had used a standard feedforward backpropagation network. Plaut et al. (1996) demonstrated that, with the added dimension of time and the accompanying notion of graded settling, an attractor network better simulated human performance in the naming task. However, because these researchers were primarily concerned with capturing phenomena associated with word naming, their models did not include a semantic module. In contrast, it is the computation of word meaning from word form that is central to the present article.

In one of the initial connectionist accounts of semantic memory, Hinton and Shallice (1991) used an attractor network to simulate deep dyslexia, a type of acquired reading disorder in which the most salient symptom is the occurrence of semantic errors in reading aloud (e.g., reading *symphony* as *orchestra*). Their theory was further developed in a series of simulations reported by Plaut and Shallice (1993), extending the investigation to other aspects of deep dyslexia and demonstrating that the general principles central to these simulations could be implemented in various network architectures. Attractor networks have also been used to investigate selective impairment of knowledge of living things versus artifacts (objects made by humans) after stroke or herpes encephalitis (Farah & McClelland, 1991), the time course of selective impairment of living things versus artifacts as a function of the progression of Alzheimer's dementia (Devlin, Gonnerman, Andersen, & Seidenberg, 1998), the effects of context on lexical ambiguity resolution (Kawamoto, 1993), and the role of feature correlations in computing word meaning (McRae et al., 1997). In this article, we focus on attractor network accounts of a primary semantic memory behavioral phenomenon, semantic priming.

Semantic Priming

Semantic priming research has played an important role in developing and testing theories of semantic memory. Semantic priming refers to the robust finding that subjects typically respond more quickly to a target word such as *doctor* when it is preceded by a semantically related word such as *nurse* versus an unrelated word such as *turkey* (for an extensive review, see Neely, 1991). Response latency is most often measured as the time it takes to name or make a lexical decision (i.e., a word/nonword decision) to the target. Research into semantic priming has been extensive because many researchers consider priming effects to directly reflect the organization of semantic memory.

Results of semantic priming experiments have often been interpreted in terms of spreading-activation theory (Anderson, 1983; Collins & Loftus, 1975; McNamara, 1992). In this account, recognizing a word involves activating its corresponding node in a hierarchically structured semantic network. Response latency in a task such as lexical decision is assumed to be related directly to the time required for a word's node to reach an activation threshold. In addition, activation spreads from the activated node to all the nodes linked to it. The existence and strength of these links are assumed to reflect learned semantic relationships between pairs of words. Thus, if *nurse* precedes *doctor* in a priming experiment, response latency to *doctor* is facilitated because its node becomes partially activated when *nurse* is read or heard. We return to spreading-activation theories of semantic processing in Experiment 2.

In contrast to spreading-activation theories, compound-cue theories propose that memory is accessed via a cue that consists of the current stimulus and its context (Dosher & Rosedale, 1989; Ratcliff & McKoon, 1988). In a single-word priming experiment, the target is the current stimulus, and the prime is its context. A central assumption of compound-cue theory is that response latency is related to cue familiarity. Thus, priming occurs because familiarity depends on the strength of the semantic relationship between the prime and the target. Because the experiments and simulations presented herein do not distinguish between compound-cue theory and an attractor network account, we do not discuss this theory further.

The most recent accounts of semantic priming have been based on attractor networks (Becker et al., 1997; Masson, 1995; McRae et al., 1997; Plaut, 1995). For example, Becker et al. (1997) used a deterministic Boltzman machine to predict the existence of long-term semantic priming (i.e., where the prime and the target are separated by several words). The key to their simulations lay in the network's incremental learning mechanism (i.e., deepening of attractor basins). Predictions from their model were borne out qualitatively by subsequent behavioral experiments. In contrast, short-term semantic priming is thought to be due to residual activation of the prime's meaning (Masson, 1995). To simulate an experiment in which a prime directly precedes its target and where both are presented visually, the orthographic pattern corresponding to the prime is input to the network so that its meaning (and possibly its phonological representation as well) is computed. With the network in a state representing the prime (e.g., *nurse*), the orthographic representation of the target (*doctor*) is given as input. Facilitation results because the distributed semantic representations of the related prime and target overlap so that some proportion of the semantic units begin in their correct states when the target is being computed. This is not the case when the prime is unrelated to the target. This method of simulating priming is particularly easy to visualize if the network's semantic units represent features or microfeatures, as they typically do. In some networks, such as the one presented herein, semantic units correspond to features that can be verbalized, such as <has wings> or <used for carpentry>, although it does not necessarily have to be so.

In understanding semantic memory, one important issue that has only recently drawn attention in the priming literature is the role of various types of semantic and associative relationships (Hodgson, 1991; Moss, Ostrin, Tyler, & Marslen-Wilson, 1995). Although many types of relation exist, the only distinction that typically is drawn is a broad one between semantic similarity and associative relatedness. Two lexical concepts are viewed as semantically similar if they share features (Shelton & Martin, 1992) or have a common superordinate category (Lupker, 1984). In contrast, two words are viewed as associatively related if one is produced as a response to the other in word association norms such as those of Postman and Keppel (1970). Because associative relatedness has been defined in this manner, it almost certainly encompasses a number of types of semantic relationships (Hodgson, 1991; McKoon & Ratcliff, 1992). As such, models of semantic memory have used various methods to capture these relationships.

Masson (1995) used an attractor network to investigate associative-based priming, which he viewed as reflecting the degree to which two words appear in the same contexts (akin to Lund & Burgess, 1996). In Masson's (1995) (Hopfield, 1982, 1984) network, associative relatedness was represented in terms of overlap of distributed word meaning. With this scheme, Masson (1995) simulated both the strong associative priming found when a prime directly precedes its target and the weak priming found when an unrelated stimulus is interleaved (Joordens & Besner, 1992).

In contrast to Masson's (1995) approach, Plaut (1995) represented both associativity and semantic similarity in his model. Associativity was operationalized as the probability of the co-occurrence of two words during the training regime, whereas semantic similarity was represented as pattern (featural) overlap. Plaut (1995) used his model to simulate priming based on both associative relatedness and semantic similarity. In addition, he simulated associative priming across an interleaved stimulus, increased priming for high- versus low-dominance category exemplars (Schwanenflugel & Rey, 1986), and increased facilitation for degraded (Becker & Killion, 1977) and low frequency targets (Becker, 1979).

The empirical goal of the modeling presented herein is most closely related to Masson (1995) and Plaut (1995) in that the aim is to capture human performance in priming tasks. However, we focused strictly on semantic-similarity priming in the absence of normative association because this type of semantic relationship is well defined, making it a suitable candidate for modeling. The ways in which other types of semantic and associative relationships might be delineated and modeled is outlined in the General Discussion.

The model itself is most closely related to those of Hinton and Shallice (1991) and Plaut (1995) in that it simulates a word form (orthography or phonology) to meaning computation by using a network trained via the backpropagation-through-time learning algorithm. The most important way in which our model improves upon previous ones is the manner in which word meaning is represented. Masson's (1995) network included a small abstract vocabulary of three pairs of associatively related items. Related words were constructed so that they overlapped on 54 of 80 semantic units. Unfortunately, it is unclear if this degree of overlap is representative of the associated prime-target pairs used in priming experiments. Plaut (1995) likewise used an abstract representation of word meaning. In his scheme, degree of semantic similarity was determined by constructing category prototypes and then probabilistically altering them to varying degrees to form within-category exemplars. Hinton and Shallice (1991) used a somewhat different method in that they hand crafted semantic feature representations for a set of artifact and living thing concepts [note that Becker et al. (1997) used Hinton & Shallice's (1991) representations]. Critically, in all of these cases, decisions regarding the appropriate amount of pattern overlap were left to the researchers themselves. In contrast, our model uses an empirically derived representation of word meaning that is based on the semantic feature production norms of McRae et al. (1997). These norms include 190 object concepts, such as *dog* and *chair,* taken from six common artifact and four living thing superordinate categories. Thus, degree of featural overlap was determined by the subjects in McRae et al.'s (1997) norming experiment, not by the researchers. Basing a model on empirically derived rather than researcher-determined semantic representations is a principled way to reduce the degrees of freedom when simulating human performance. This is relevant in the present case because MB demonstrated that the amount of semantic-similarity priming depends crually on the degree of prime-target featural overlap. Thus, an independent measure of semantic overlap is critical for demonstrating that an attractor network produces quantitatively appropriate semantic-similarity priming effects.

Semantic-Similarity Priming without Association

The experiments and modeling focused on featurally similar items, such as *goose-turkey*, for which subjects do not respond with one member of the pair when presented with the other in a word association task. One additional constraint was imposed by the fact that our model includes no mechanism to simulate subjects' strategies. Therefore, we focused on priming experiments that were designed to minimize strategic processing. Following a number of researchers, we assumed that a short time interval between the presentation of the prime and the target (stimulus onset asynchrony, SOA) minimizes subjects' strategies. de Groot (1984) and den Heyer, Briand, and Dannenbring (1983), among others, have shown that strategic effects are minimal when the SOA is 250 ms or less. In addition, it has recently been argued that presenting items in a sequence without overtly pairing primes and targets also minimizes strategic effects (McNamara & Altarriba, 1988; Shelton & Martin, 1992). Both methods appear to be most effective when a low proportion of the items are related. An experiment in which strategic effects are minimized is often referred to as tapping automatic, lexical-internal mechanisms (Neely, 1977; Posner & Snyder, 1975).

Behavioral investigations of semantic-similarity priming in the absence of normative association have produced mixed results. Chiarello, Burgess, Richards, and Pollock (1990) found semantic-similarity priming when using a 575-ms SOA with both lexical decision and naming tasks. Although this SOA is quite long, the relatedness proportion was low, and no evidence of strategic processing was found. Moss et al. (1995, Experiment 1) also found priming for semantically similar prime-target pairs when using a paired presentation auditory lexical-decision task. However, the SOA was substantially longer than 250 ms because it comprised the time required for the auditory prime to unfold plus a 200-ms interstimulus interval. Because no checks for strategic processing were included, Moss et al. (1995) were concerned that the effects may have been inflated. Therefore, their second experiment featured a single-presentation technique in which there was a constant 1000 ms between the offset of one auditory stimulus and the onset of the next. Priming was obtained in some conditions, but it was greatly reduced. Finally, Fischler (1977) obtained a large effect of semantic similarity; however, primes and targets were presented simultaneously in his experiment and subjects made a double lexical decision ("Are both letter strings words?"). Because this task encourages subjects to evaluate prime-target relationships, the priming effects probably include a substantial strategic component.

Four experiments have failed to find automatic semantic-similarity priming. Lupker (1984, Experiment 3) used a 250-ms SOA with a naming task and found a small (6 ms) priming effect. Moss et al. (1995) presented primes and targets visually in their Experiment 3 (as opposed to auditorily in their Experiments 1 and 2) and found no similarity-based priming when using single presentation with an intertrial interval (ITI) of 500 ms and a lexical-decision task. Likewise, Shelton and Martin (1992) found no priming when using a single-presentation, lexical-decision task with both a 500-ms (Experiment 3) and a 200-ms ITI (Experiment 4).

These empirical inconsistencies were resolved by MB. Their experiments were based on the observation that the items used by Shelton and Martin (1992) and Moss et al. (1995) were not highly similar (see also Lund, Burgess, & Atchley, 1995). This was presumably true of Lupker (1984) as well, who created items by randomly pairing exemplars from within a superordinate category, such as *clothing*. To address this problem, MB constructed items that were more similar than those of the previous studies, as established by subjects' ratings. Word association norms showed that the items were not associated in either the prime-target or target-prime direction. In MB Experiment 1, priming was found in four conditions formed by crossing presentation technique (250-ms SOA paired presentation versus single presentation with a 200-ms ITI) and task (lexical decision versus a "Does it refer to a concrete object?" semantic decision). They further illustrated the role of degree of similarity in their Experiment 3 by using triplets of lexical concepts in which a target, such as *beans*, was paired with both a highly similar prime, such as *peas*, and a less similar prime, such as *garlic*. With a 250-ms SOA and a concrete object-decision task, highly similar prime-target pairs showed robust priming, whereas less similar pairs did not. In addition, decision latencies were shorter when targets were preceded by highly similar versus less similar primes. A somewhat different pattern of results obtained with a 750-ms SOA. Although the difference between highly similar and less similar items remained, they both produced priming compared to the dissimilar condition. Finally, MB also suggested that Shelton and Martin (1992) and Moss et al. (1995) obtained null results partly because they used targets that were shorter in length and higher in Kucera and Francis (1967) frequency than those of MB, thus providing less opportunity to observe target facilitation.

Overview

The remainder of this article is structured as follows. We first describe the attractor network. After that, simulations of MB Experiments 1 and 3 are presented. Note that the primes and targets used in these experiments are a subset of the items trained in the model. Two predictions are then presented, and human experiments testing them are presented. The first prediction was that semantic-similarity priming in the absence of normative association should be equivalent when the primes and targets are presented in reversed order. Although researchers, such as Plaut (1995), have made this prediction, it has yet to be tested in models or humans. We then tested a prediction that runs counter to a central aspect of hierarchically based spreading-activation theories. In experiments based on semantic network theory, primes and targets have typically been treated as semantically related if they are exemplars of the same superordinate category; for example, *beans* and *peas* are both vegetables (Moss et al., 1995). Priming is expected because activation is assumed to spread from the prime to the target via links with their shared category node

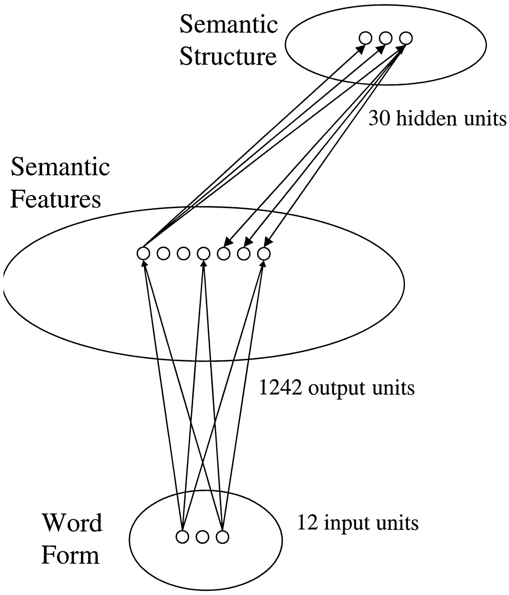


Figure 1. Architecture of the attractor network. Not all connections are shown. Where connections are shown, full connectivity was used.

(Collins & Loftus, 1975). However, an attractor network such as ours does not contain superordinate category nodes, so that priming must be due to featural overlap epiphenomenon. These predictions were contrasted by using triplets in which, for example, *squash* was the target, *pumpkin* was the more similar/less typical prime, and *corn* was the less similar/more typical prime. Supporting an attractor network account, priming was obtained only when the target was preceded by its more similar/less typical prime.

II. THE MODEL

In this part of the article, the architecture of the model is described, along with details of the input and output representations. The manner in which word meaning is computed in the network and the training regime are also presented.

Network Architecture

The model's architecture is presented in Figure 1. The network mapped directly from 12 word form units (representing orthography or phonology) to 1242 semantic feature units (word meaning). In addition, the semantic feature units looped back to themselves through a layer of 30 hidden units that encoded higher-order feature intercorrelations; hence, they are called semantic-structure units. All connections were unidirectional.

Word Form Units. The input to the network was an abstract word form representation that could be interpreted as either spelling or sound. Each word's form was represented by turning on 3 of the 12 units (activation = 1) with the remaining 9 being off (activation = 0). All words had a unique word form. Thus, 190 of the possible 220 3-unit patterns were used, resulting in highly overlapping representations. This scheme was designed to capture two characteristics of the mapping from English spelling or sound to meaning. First, the mapping from form to meaning is largely arbitrary for monomorphemic English words such as the 190 artifact and living thing concepts encoded in our network. This means that subword components, such as *c,* do not map directly to specific components of meaning. Second, the representation of words over letters and phonemes is relatively dense when the language as a whole is taken into account. Both of these principles were instantiated in this representation in that each unit participated in numerous words (M = 48 words, *range* = 43–51).

Semantic Units. Semantic representations for the 190 lexical concepts were taken from McRae et al. (1997). In their norms, subjects were given the names of exemplars, such as *dog* and *chair,* and were asked to list features. There were 19 exemplars from 10 object categories: birds, mammals, fruit, vegetables, clothing, furniture, kitchen items, tools, vehicles, and weapons. McRae et al. (1997) retained all features that were listed for a concept by at least five of 30 subjects, resulting in a set of 1242 features, such as <has a beak> and <used by children>. Thus, there were 1242 semantic units in the network, each representing a feature from the norms. Lexical concepts were represented as distributed binary patterns over the semantic units; feature salience was not coded. The resulting representation was sparse because a concept consisted of at most 27 of the 1242 features. Note that representing word meaning in this manner is not intended as a theoretical statement concerning the precise way that information is stored in people's semantic memory. Verbally produced semantic features should be considered as a window into people's knowledge of lexical concepts, rather than an accurate portrayal. On the other hand, semantic features such as these are a valuable resource in that they provide a representation of semantic information that is interpretable and can be used as the basis for both computational modeling and human experimentation.

Semantic-Structure Units. Forster (1994) stated that because the mapping from word form to meaning is largely arbitrary, and because connectionist networks are not particularly adept at learning arbitrary mappings, they are inappropriate for simulating lexical conceptual processing. However, McRae et al. (1997) noted that relevant structure is not restricted to mapping between domains, such as orthography, phonology, and semantics, but also includes structure within a domain. To show that this structure existed in the semantic representations derived from their norms, they computed pair-wise feature correlations and found over 1000 significantly correlated feature pairs, such as <has wings> and <has a beak>. Using feature verification and semantic priming tasks, McRae et al. (1997) provided evidence that people encode semantic structure in the form of feature correlations and that this knowledge plays a key role in computing word meaning. To simulate this aspect of the computation, 30 semantic-structure units received input from the semantic layer and fed directly back to it. The role of these hidden units (labeled "clean-up units" by Hinton & Shallice, 1991) was to encode semantic regularities and to exploit them when computing word meaning.

To gain insight into the network's processing dynamics, it is helpful to view the semantic loop as being somewhat like an autoencoder that compensates for the deficiencies in the direct word form to meaning mapping. An autoencoder performs a computation akin to principle components analysis by compressing data at the hidden unit layer and then restoring it at the output layer (Cottrell, Munro, & Zipser, 1987). It differs from principle components analysis in that the variance accounted for by each factor is spread relatively equally across the hidden units (semantic-structure units, in the present case), and there is no orthogonality constraint. Thus, this part of the network encoded the higher-order regularities in the featural representations. Note that because the semantic loop was fully unidirectionally connected, the network could, theoretically, encode regularities involving any number of features. This differs from the Hopfield (1982 and 1984) network used by McRae et al. (1997) in which pair-wise feature correlations were encoded in weights directly connecting semantic units.

Computing the Meaning of a Word. To compute a word's meaning, its form was hard-clamped at the input layer (i.e., it was provided at each time tick). Net input to the semantic and semantic-structure units was calculated by summing the activations from all sending units times the connecting weights. Note that activation was free to pass between all layers at each time tick. Therefore, the semantic and semantic-structure units were initialized to random starting values in the range .20 ± .05 during training. Each tick of processing time allowed activation to spread one layer forward. Activation was passed between layers for 20 ticks of time, segmented into four time steps, each consisting of five time ticks. The number of time steps is relatively arbitrary, and the total number of time ticks determines the weighting proportion of the net input averaging. Dividing computation time into discrete ticks enables us to approximate a continuous system on a digital computer (for discussion, see Plaut et al., 1996). Net inputs x_j were averaged according to Equation 1:

$$x_j^{[t]} = \tau \sum_i {}_i^{[t-\tau]} w_{ji} + (1 - \tau) x_j^{[t-\tau]} \tag{1}$$

where s_i is the activation of *unit*$_i$, and w_{ji} is the weight on the connection to *unit*$_j$ from *unit*$_i$. Therefore, a unit's input at each time step was a weighted average of its current input and the input from all sending units. Activation (s_j)was determined by the sigmoidal function,

$${}_j^{[t]} = \frac{1}{1 + \exp(-x_j^{[t]})} \tag{2}$$

where x_j is the net input to *unit*$_j$, and exp(.) is the exponential function.

Learning. Error was backpropagated over the 20 processing ticks in a manner analogous to the forward pass. Error was injected into the system (i.e., the target semantic representation was provided) for the final two steps (10 time ticks). Because it was not injected for the initial two, the network was allowed to produce the target output gradually over time. Error derivatives were calculated by using cross-entropy, as opposed to the more typical sum squared error. Cross-entropy is preferred in this case because it is appropriate to consider the semantic features as binary, being either present (1) or absent (0). Thus, states between on and off can be viewed as representing the probability that the feature is present or absent (Hinton, 1989; Rumelhart, Durbin, Golden, & Chauvin, 1995). Cross-entropy error (E) summed over the p patterns was calculated as,

$$E = \tau \frac{\sum_{t=10}^{19} \sum_{j} d_j \ln (s_j) + (1 - d_j) \ln (1 - s_j)}{10} \qquad (3)$$

where d_{j} is the desired activation for $unit_{\mathrm{j}}$, and s_{j} is that unit's computed activation. Note that for the backpropagation-through-time algorithm, error was averaged over the final 10 time ticks ($10 \leq t \leq 19$) and then scaled by τ. Cross-entropy produces large error values for units that are on the wrong side of .5 (off when they should be on, or vice versa), and small error values for units on the right side of .5, thus pushing units to the correct half of the probability scale.

Weight changes were calculated by using the backpropagation-through-time algorithm. This is computationally equivalent to feedforward backpropagation, where the network is unfolded over time ticks to produce a series of identical networks that are connected by the same weights. Error derivatives were calculated between sending units at time tick t-1 and receiving units at time t. No restrictions were placed on the types of weights that could be formed. Soft-clamping of semantic units was used to accelerate learning, which means that when an exemplar was processed, the correct semantic pattern was injected as a small addition to the net input (a value of 1 was injected to each unit that should be on) to facilitate development of semantic attractors. This process, in addition to using cross-entropy, helped the sparse network learn to turn units on (a reasonable solution for minimizing error in a large, sparse network would be to turn off all output units). Later in training, when larger weights had been formed, injection essentially had no influence on processing because it was minimal compared to the internal net input. Soft clamping of output units was, of course, turned off for all testing runs.

The network was trained by using the PDP++ (Version 1.1) simulator developed at Carnegie Melon by R. C. O'Reilly, C. K. Dawson, and J. L. McClelland. Weights were updated after each pattern presentation. The learning rate was .01 throughout training. Momentum was set at 0 for the first 10 epochs of training and .9 thereafter. Each epoch consisted of randomly presenting the 190 patterns. Thus, all concepts were equally frequent for the network. After 85 epochs of training, mean cross-entropy per pattern over the 1242 semantic units was 0.84 (SE = 0.08). Note that this statistic was calculated only at the final time tick and so differed from Equation 3 in that it was not averaged over ticks

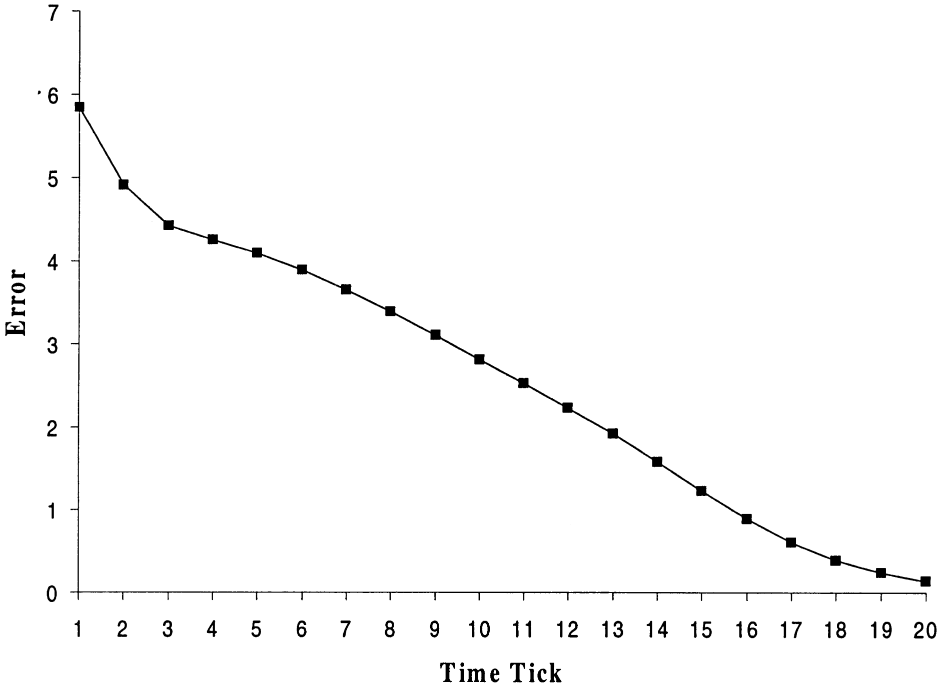


Figure 2. Settling profile as averaged over the 190 trained concepts.

nor was it scaled by τ. As an indication to the degree to which the model was trained, the number of units on the wrong side of .3 (for units that should be off) and .7 (for units that should be on) was calculated, summed over all patterns. Of the 233,017 units that should have been off, 15 (0.006%) were activated above .3. Of the 2,963 units that should have been on, 9 (0.3%) were activated below .7. Figure 2 shows the mean settling profile over 20 ticks for the 190 lexical concepts. These values were calculated by averaging cross-entropy at the semantic layer across five runs in which a word's form was clamped after the computation of a dissimilar concept. Thus, in this simulation, the starting point for a pattern was not random; rather, the semantic and semantic-structure units were in the state determined by computing a dissimilar concept for 20 ticks.

The role of the semantic-structure units can be illustrated by analyzing the settling patterns of the concepts. McRae et al. (1997) showed that the 190 normed concepts varied in terms of the degree to which their features were intercorrelated. One way to operationalize this notion is to compute intercorrelational density for a concept. To do this, we calculated the Pearson product moment correlation for each pair of features that appeared in at least three concepts (to avoid spurious correlations). Intercorrelational density was measured by summing the percentage of shared variance between each pair of a concept's features that were correlated at the $p < .01$ level. For example, the intercorrelational density of *grapefruit* was 2009 (many of its features are correlated with one another), whereas it was just 9 for *carpet*. This measure can be contrasted with one based on

TABLE 1
Using the Number of Individual Features and Intercorrelational Density of a Concept to Predict Cross-entropy as the 190 Concepts Settled

	Number of individual features			Intercorrelational density		
Time tick	r^2	$F(1, 188)$	$p <$	r^2	$F(1, 187)$	$p <$
1	.75	573.98	.0001	.02	3.35	.1
2	.74	536.77	.0001	.02	3.09	.1
3	.76	601.32	.0001	.03	5.10	.05
4	.75	569.22	.0001	.03	5.34	.05
5	.75	556.51	.0001	.03	5.56	.05
6	.76	578.42	.0001	.02	4.19	.05
7	.75	552.53	.0001	.01	<1	
8	.71	454.00	.0001	.01	<1	
9	.65	348.71	.0001	.04	7.13	.01
10	.57	252.97	.0001	.08	16.63	.0001
11	.49	181.58	.0001	.12	25.07	.0001
12	.41	129.77	.0001	.13	27.11	.0001
13	.32	88.81	.0001	.13	26.77	.0001
14	.24	58.27	.0001	.13	28.67	.0001
15	.16	36.18	.0001	.14	29.64	.0001
16	.11	22.03	.0001	.14	29.25	.0001
17	.06	11.24	.001	.13	27.31	.0001
18	.03	6.27	.02	.11	24.17	.0001
19	.01	1.66	.2	.10	21.76	.0001
20	.00	<1		.11	23.32	.0001

Note. The individual features factor was forced into the regression equation before intercorrelational density.

individual features, such as the number of features in each concept. The role of both of these factors in the network's dynamics was illustrated by using them to predict cross-entropy at each of the 20 time ticks in the data from the simulation presented in Figure 2. It was expected that the number of individual features would have a substantial effect on cross-entropy because this factor determines the number of on-units in the target. The importance of feature correlations was highlighted by forcing the number of individual features into the regression equation and then using intercorrelational density to predict the residual variation. Table 1 shows the influence of semantic structure over and above the number of individual features. Note that the effect of feature correlations builds over time because a number of time ticks are required for the relevant information to feed from the semantic units through the semantic-structure units and then back again.

In summary, the key elements of our model are: 1) the semantic representations are based on empirically generated features; 2) a word's meaning corresponds to a point attractor in semantic space; 3) the mapping from word form to meaning is strained because of overlap in the formal representations as well as the lack of systematicity in the mapping between the domains; 4) semantic regularities play a role in processing, helping to overcome this deficiency in the mapping; and 5) semantic representations are sparse.

The model differs in a number of ways from an earlier version that was reported by McRae et al. (1997). They were primarily interested in the role of feature correlations in the computation of word meaning, and so they used a Hopfield (1982 and 1984) network that enabled transparent encoding and analysis of the correlational information. One disadvantage of the Hebb (1949) learning rule used for training their model is that it is computationally weak, as evidenced by the fact that McRae et al. (1997) were able to store only about 80 of the 190 concepts. Therefore, the present model used a more robust learning algorithm, backpropagation-through-time, that enabled all 190 concepts to be learned. Incorporating a larger sample of items allows a better approximation of semantic structure to be encoded in the network and makes it possible to include all concepts used in MB's experiments. A further aspect of using the present architecture is that the semantic-structure units, as opposed to direct correlational connections between feature pairs, enabled the model to capture higher-order statistical relationships among features. Third, the elongated time scale (i.e., the present model used a greater number of time ticks for a concept to converge) provides a better opportunity to observe subtle differences among settling times when simulating decision latencies. In addition, it is advantageous to show that theoretical principles such as those outlined in the previous paragraph can be implemented in multiple architectures.

Finally, it should be noted that our model is not an accurate simulation of how children learn lexical concepts. Research in this area points to complex interactions between linguistic and non-linguistic input while children learn word meaning (e.g., Barrett, Abdi, Murphy, & Gallagher, 1993; Gelman, 1988; Jones & Smith, 1993). Because of the simplifications in the way the network was trained, it is not appropriate to use it to simulate developmental data. However, the network was designed to capture facts concerning how adults perform on speeded lexical tasks, and it is the end point of learning that is of interest here. We now turn to the simulations of recent behavioral investigations of semantic-similarity priming.

III. SIMULATION 1: MCRAE AND BOISVERT (1998, EXPERIMENT 1)

MB Experiment 1 demonstrated semantic-similarity priming effects with items that were highly similar and not normatively associated (see Appendix A for their items, but with the primes and targets reversed). Priming was found in four conditions: 250-ms SOA paired presentation with semantic and lexical decision tasks and single presentation with a 200-ms ITI and the same two tasks. We chose to simulate this experiment because all of the items used in it were included in the model's training set, and the priming effects were robust in all four conditions. Note that the manner in which priming was simulated in this article (and in Masson, 1995; Plaut, 1995) best resembles short SOA paired presentation of prime and target where subjects silently read the prime and then respond to the target. Therefore, all experiments simulated in this article used short SOA paired presentation. In terms of task, we believe that the simulations best correspond to a semantic task such as "does the word refer to a concrete object?" because semantic settling time serves as a proxy for decision latency (for discussions of the empirical advantages of using semantic decisions to study the computation of word meaning, see MB; McRae et al., 1997). Semantic settling time may also be used to simulate lexical decision latencies to the extent that human lexical decisions are influenced by the computation of word meaning (for a discussion of this issue, see Joordens & Becker, 1997). Finally, there are two reasons why it would be unwise to compare the results of naming experiments with our simulations. First, the model does not involve the computation of phonology from orthographic input. Second, the extent to which word meaning influences the computation of phonology depends on the interaction of a number of factors, so it is not straightforward to map from semantic convergence latencies to naming latencies (Strain, Patterson, & Seidenberg, 1995). In summary, it was predicted that the network would settle faster for targets that were preceded by highly similar versus dissimilar primes, as was observed in humans by MB.

Method

A priming trial was simulated as follows. Before presenting the prime, all semantic and semantic-structure units remained in the state determined by the previous target.[1] The word form representation of the prime was hard-clamped for 15 ticks.[2] The mean prime cross-entropy at the final prime tick is shown as the first point in Figure 3 (upper panel). After the prime settled for 15 ticks, the target's word form was clamped with all other units unchanged. The target was allowed to settle for 20 ticks, and cross-entropy at the semantic layer was recorded. Simulating MB Experiment 1 consisted of conducting priming trials for five random orders of their 38 similar and 38 dissimilar prime-target pairs.

Results

Figure 3 (upper panel) shows the target settling profiles. The priming effect is clear in that the similar targets have a lower cross-entropy at each of the 20 time ticks. Note that the cross-entropy of the primes after 15 prime ticks are equivalent for the similar and dissimilar conditions because the dissimilar primes in MB Experiment 1 were formed by re-pairing the similar primes.

To conduct statistical analyses on the network's performance, it was necessary to identify an appropriate measure to serve as an analogy to subjects' decision latencies. The most typical method to derive this estimate from an attractor network is to measure the number of cycles required to reach a stable state. However, it is probable that people initiate their response before reaching the functional equivalent of a stable state, particularly in a task in which both speed and accuracy are stressed. Furthermore, the precise location of this threshold is unclear, and so it is advantageous to present results at a number of points. Therefore, we used the number of ticks required for a target to reach certain levels of cross-entropy (referred to as convergence latency throughout the article). Nine levels, ranging from 2.5 to 0.5 in decrements of 0.25, were sampled. These were chosen by inspecting Figure 2 and were designed to span a reasonable range of possible values.

A related issue concerns determining whether the model is making a correct response. This issue is difficult to resolve because it is problematic to determine whether a computed representation is correct, or more precisely, whether it would support a correct response

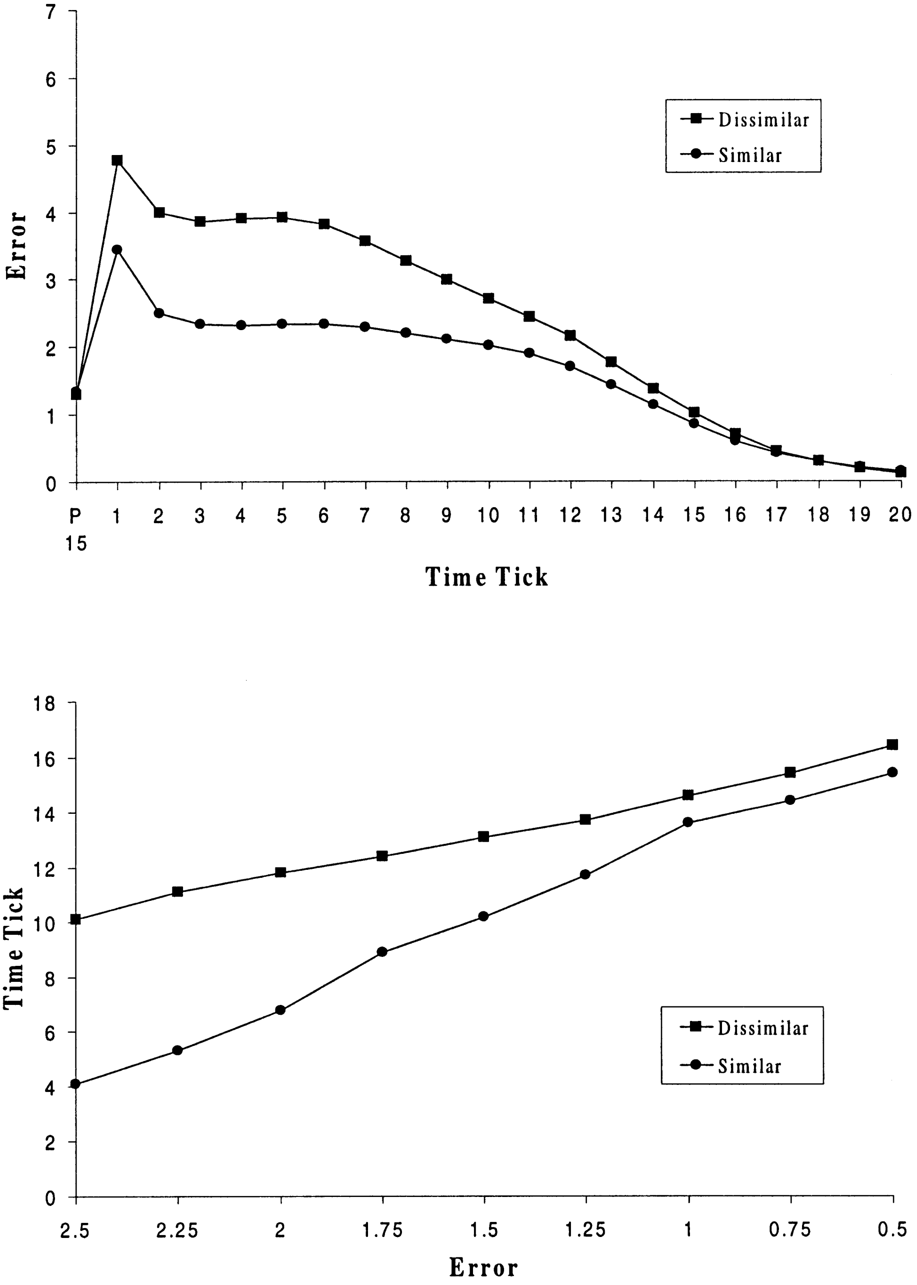


Figure 3. Upper panel: The settling profiles averaged over the 38 similar and dissimilar targets from MB Experiment 1. Lower panel: The number of ticks required to reach various levels of cross-entropy.

for the task at hand. For example, it could be considered as correct when all units are on the right side of .5, or when a relatively strict criterion, such as all units being less than .05 when they should be off and greater than .95 when they should be on, is satisfied. Various intermediate criteria are also possible. Given these indeterminacies, our approach was to reduce the number of free parameters in the simulations by making the minimal number of assumptions and providing views of the simulation data spread across time. Therefore, we did not define a model error, so that all items are included in the simulation analyses presented herein.

Figure 3 (lower panel) presents the mean number of ticks required to reach each level of cross-entropy. A two-way repeated measures analysis of variance (ANOVA) was conducted with error level (2.5–0.5, decrements of 0.25) and prime-target similarity (similar versus dissimilar) as the independent variables and number of ticks to reach the specified error level (convergence latency) as the dependent variable. Prime-target similarity influenced convergence latency, $F(1, 37) = 41.94$.[3] In addition, similarity and error level interacted because priming effects in the network decreased as error level approached 0.5, $F(8, 296) = 16.44$. Planned comparisons revealed significant priming effects at the first eight error levels, with 0.5 being the exception. Finally, convergence latency increased across the nine error levels, $F(8, 296) = 168.26$.

Discussion

The network clearly produces priming effects that are analogous to humans for items that are semantically similar but not normatively associated. The priming effects resulted because the similar primes and targets overlap to a greater degree than the dissimilar ones at the semantic and semantic-structure layers. We quantified this overlap by computing the cosine between the 1242-unit semantic feature vectors for the primes and targets and between the 30-unit semantic structure vectors by using the states of these units after 20 time steps. (Note that the states of the semantic-structure units were recorded when the concepts were presented for five random orders, rather than during the priming simulations themselves.) The mean cosine for the semantic feature representations was .51 (SE = .03) for the similar prime-target pairs and .02 (SE = .01) for the dissimilar pairs, $t(37) = 15.04$. When computed across the semantic-structure units, the mean cosine was .74 (SE = .02) for the similar prime-target pairs and .54 (SE = .01) for the dissimilar pairs, $t(37) = 10.13$. Thus, the network is not simply a reiteration of the information contained in the feature norms because it combined the influence of both individual features and the correlations among them.

As in Tversky's (1977) contrast model, three aspects of featural similarity contribute to priming effects in the network. First, the features shared by the prime and the target are important because they are activated when the target is presented. Note that they may be more or less active depending on the number of ticks for which the prime has been computed, as well as the computational dynamics of the specific prime. Second, features distinctive to the target must be activated. Third, features distinctive to the prime must be turned off. The ease with which distinctive features change their states depends on a number of factors, a primary one being feature correlations. The weights to and from the semantic-structure units encode what features tend to be on simultaneously. Thus, if a feature distinctive to the target is correlated with a number of the other features of the target concept, it is easily activated. That is, although the distinctive feature has not been activated as part of the prime concept, it is easily activated because of the presence of intercorrelated features in the target, particularly if the prime contains a subset of those features (i.e., some correlated features are activated before the target's word form is presented). In contrast, if a feature is distinctive to the prime and is correlated with a number of the other features of the target concept (i.e., the target concept violates the featural regularities), it may be difficult to deactivate because it is supported by activated, correlated features. In these ways, the semantic-structure units play a key role in determining the magnitude of the priming effects (for a more detailed discussion of the role of feature correlations in determining priming effects, see McRae et al., 1997).

The results of Simulation 1 differed from those of the semantic-similarity simulation presented in Plaut (1995). At the time that Plaut conducted his simulations, Shelton and Martin (1992) had reported experiments that suggested that little or no semantic-similarity priming existed with short SOAs, in contrast to associative relationships that showed robust priming. Thus, Plaut (1995) constructed prime-target pairs by randomly sampling two exemplars from within his abstract superordinate categories, as was done in Lupker (1984). This approach yielded small priming effects, much like those obtained by Lupker (1984) and Shelton and Martin (1992). However, Plaut's (1995) network would presumably show strong priming with highly similar items. In fact, large priming effects were obtained when a prototype was used to prime a similar (high dominance) exemplar.

In summary, the attractor network demonstrated an appropriate magnitude of priming when the degree of conceptual similarity was determined by the subjects of McRae et al.'s (1997) feature norming experiment.

IV. SIMULATION 2: MCRAE AND BOISVERT (1998, EXPERIMENT 3)

Having demonstrated a semantic-similarity priming effect, we now turn to a somewhat more subtle demonstration. MB explained the difference between their Experiment 1 and the null effects of Moss et al. (1995) and Shelton and Martin (1992) primarily in terms of differing degrees of prime-target similarity. To provide evidence for this, they designed word triplets in which a target (*jar*) was paired with both a highly similar (*bottle*) and a less similar (*plate*) prime, with the degree of similarity being established by subjects' ratings. This manipulation enabled the investigation of the effects of semantic similarity while the characteristics of the target were held constant. The highly similar items were a subset of those used in MB Experiment 1, and the prime-target similarity of the less similar primes was in the range of Shelton and Martin's (1992) items. MB Experiment 3 also incorporated SOAs of 250 ms and 750 ms because most researchers believe that priming effects at a short SOA, such as 250 ms, reflect lexical-internal factors, whereas priming effects at a longer SOA, such as 750 ms, may be influenced by subjects' strategies (de Groot, 1984; den Heyer et al., 1983; Neely, 1977). Consistent with these notions, with a 250-ms SOA, semantic decisions ("Does it refer to a concrete object?) were faster for targets that were preceded by highly similar primes (685 ms) than for those preceded by less similar (712 ms) or dissimilar primes (711 ms), and no priming obtained for the less similar items. With a 750-ms SOA, semantic decisions were again faster in the highly similar condition (646 ms) than in either the less similar (664 ms) or dissimilar conditions (692 ms). In addition, there was reliable priming for the less similar items, replicating Shelton and Martin's (1992) finding of priming for their items with the same SOA.

Simulation 2 investigated whether the network would exhibit settling profiles that were appropriate for these three groups of items. That is, the model should show faster settling times for targets preceded by highly similar primes versus either less similar or dissimilar primes. The prediction for less similar versus dissimilar primes was derived as follows. In the human data, the less similar items showed no priming at the short SOA, but significant priming at the long SOA. Because we assume that the simulation best approximates short SOA priming, the difference in convergence latency for the less similar and dissimilar conditions should be small. Note that we did not attempt to approximate closely the long SOA condition because it is unclear how to incorporate subjects' strategies into the network. Regardless, less similar targets should converge somewhat faster than dissimilar ones because there must be a basis for the priming obtained at the long SOA (i.e., for this difference to be observed, these groups must differ to some extent). Therefore, a small difference between these conditions should be observed.

Method

The method was identical to Simulation 1 except that the items were the 27 word triplets from MB Experiment 3. See MB Appendix B for those items.

Results

The settling profiles for the targets preceded by each type of prime are presented in Figure 4 (upper panel). The difference between the highly similar and less similar pairs is more pronounced than between the less similar and dissimilar pairs, reflecting the human data.

Figure 4 (lower panel) presents the mean number of time ticks required to reach each level of cross-entropy. As in Simulation 1, a two-way repeated measures ANOVA was conducted with error level (2.5–0.5, decrements of 0.25) and prime-target similarity (highly similar versus less similar versus dissimilar) as the independent variables, and number of ticks to reach the specified error level (convergence latency) as the dependent variable. Prime-target similarity influenced convergence latency, $F(2, 52) = 26.75$. In planned comparisons conducted on the main effect of similarity (i.e., when the nine error levels were combined), convergence latency for the highly similar items was significantly shorter than for the less similar, $F(1, 52) = 30.35$, and dissimilar, $F(1, 52) = 47.90$, items, whereas the less similar and dissimilar targets did not differ to a significant degree, $F(1, 52) = 1.99$; $p > .1$. Convergence latency increased across the nine error levels, $F(8, 208) = 147.14$.

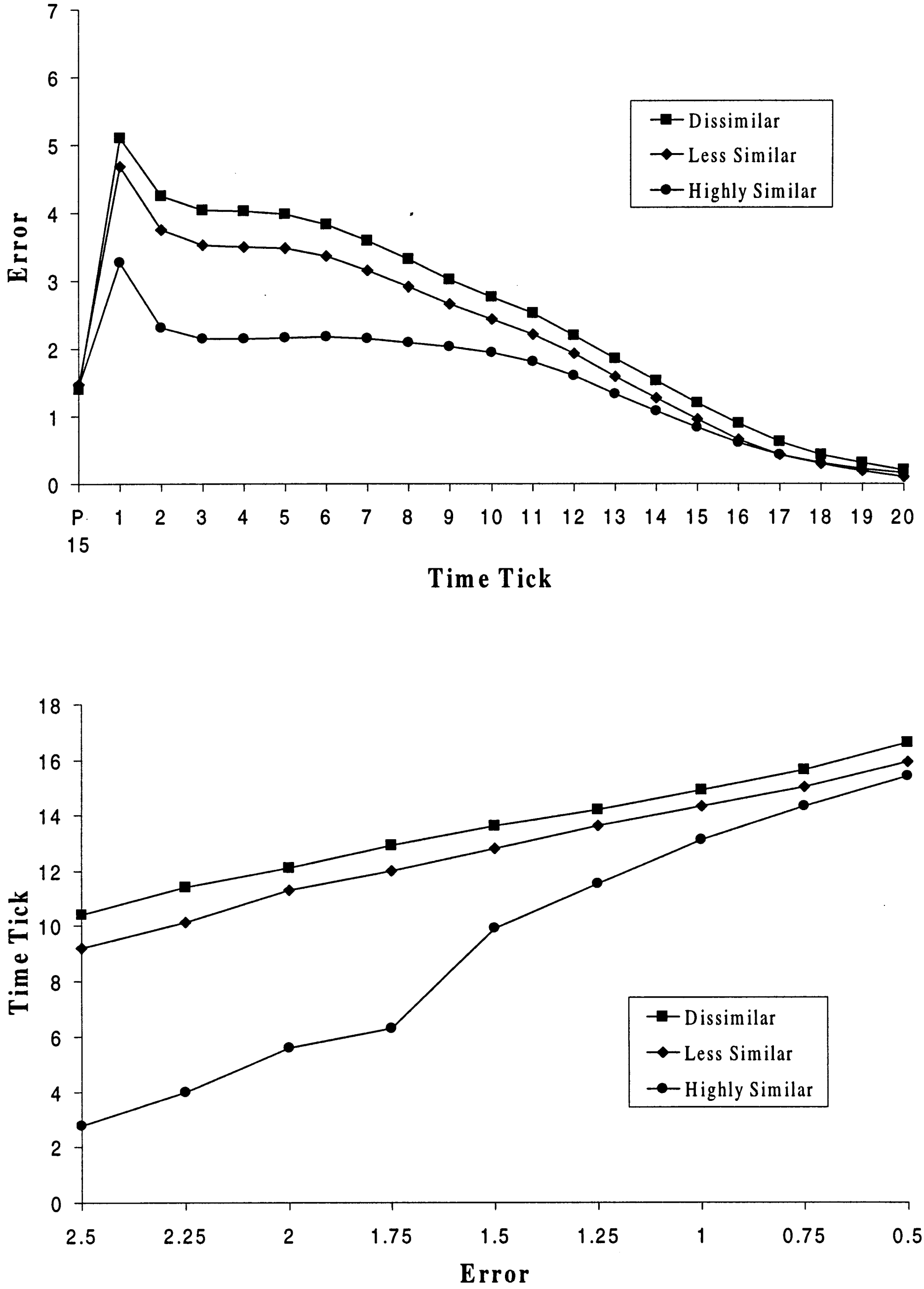


Figure 4. Upper panel: The settling profiles averaged over the 27 highly similar, less similar, and dissimilar targets from MB Experiment 3. Lower panel: The number of ticks required to reach various levels of cross-entropy.

As in Simulation 1, similarity and error level interacted because differences among the three conditions decreased at the lower error levels, $F(16, 416) = 17.75$. Planned comparisons explored the differences among the three groups at each error level. At 2.5 and 2.25, convergence latency for highly similar targets was significantly shorter than for less similar targets, which was in turn shorter than for dissimilar targets. These results mirrored those of the long SOA presentation condition of MB Experiment 3. At the next five error levels (2 to 1), highly similar targets converged more quickly than did less similar targets, which did not differ significantly from the dissimilar condition. These results mirrored those of the short SOA condition of MB Experiment 3. Finally, at error levels of 0.75 and 0.5, only the highly similar and dissimilar groups differed reliably.

Discussion

The priming effects demonstrated by the model reflect the subtle human effects found when prime-target similarity is varied at short and long SOAs. The small difference between the less similar and dissimilar conditions illustrates why a manipulation of this magnitude would result in marginal or null priming effects, as was obtained with humans by Lupker (1984), Moss et al. (1995), and Shelton and Martin (1992). Note that it is the confluence of factors, such as prime-target similarity, target length, and target frequency, that determines the probability of finding a significant priming effect in any single experiment. In fact, the prime-target pairs of both Shelton and Martin (1992) and Moss et al. (1995) were somewhat more similar than those of MB Experiment 3, although their targets were significantly higher frequency (Kucera and Francis, 1967) and shorter in length in letters (MB). Thus, intermediate prime-target similarity combined with reasonably short, high frequency targets contributed to their null effects.

This simulation is consistent with that of Plaut (1995), who found that semantic priming interacted with target category dominance. High dominance items (items for which the exemplars differed only slightly from the prime prototypes) showed a larger effect than did low dominance items (where the exemplars shared fewer features with the prime prototype). Category dominance, at least as defined in this manner, is a type of similarity relation.

V. REVERSAL OF PRIMES AND TARGETS

The purpose of Simulation 3 and the accompanying Experiment 1, was to determine whether semantic-similarity priming is equivalent when the primes and targets are presented in reversed order. This hypothesis was tested by reversing the items of MB Experiment 1. A number of studies have compared the direction in which the prime and target are presented using items that are directionally associated (Koriat, 1981; Peterson & Simpson, 1989; Seidenberg, Waters, Barnes, & Langer, 1984; Shelton & Martin, 1992). For example, Seidenberg et al. (1984) used items, such as *fruit-fly,* that are temporally associated in the forward but not the backward direction. Differential priming effects have been found in a number of conditions in the experiments listed above, although the results have varied depending on the SOA, task (naming versus lexical decision), and whether the prime appeared alone or in a sentence context.

Presentation order has not been investigated for semantically similar items that are not normatively associated. In fact, although a number of researchers (e.g., Plaut, 1995) have stated that this type of relation should produce equivalent priming in both directions, this has not been demonstrated in an implemented model nor in a human experiment. At first glance, it appears obvious that semantically similar priming effects should be directionally insensitive because shared features are identical regardless of presentation order. There are, however, three reasons why effects might differ. First, patterns of distinctive features change with order, and this factor will interact with patterns of feature intercorrelations. Because there is variation in the features that must be activated or deactivated when the target's meaning is computed, priming effects might vary. Second, target characteristics, such as word length and concept familiarity, influence priming so that priming effects could vary to the extent that they are not balanced. Note, however, that the primes and targets of MB Experiment 1 do not differ on these dimensions. Third, Tversky (1977) showed that similarity ratings may differ directionally under some circumstances. Specifically, he found that ratings differed when one concept was viewed by subjects as the anchor, as when subjects rate the similarity between China and North Korea, with China anchoring the judgment. However, the items of MB Experiment 1 were all basic level concepts, so anchoring effects are unlikely. Consistent with this, similarity ratings did not differ by direction. In summary, because the differences in presentation order were confined to the way in which distinctive features patterned, it was predicted that the simulated and human priming effects would be equivalent when the primes and targets were reversed.

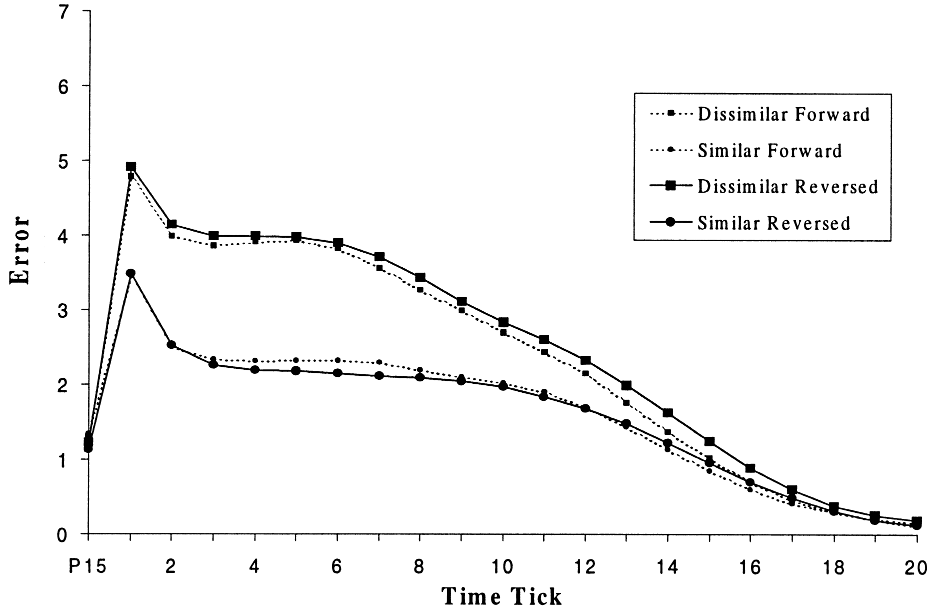
Simulation 3: McRae and Boisvert (1998, Experiment 1) Reversed

The goal of Simulation 3 was to derive a prediction for Experiment 1.

Method. The method was identical to Simulation 1 except that the similar primes now served as targets and the targets as primes. The manner in which the new primes were re-paired to form the dissimilar primes differed slightly from MB Experiment 1 (see Appendix A).

Results. Figure 5 (upper panel) shows the target settling profiles for Simulations 1 and 3. The priming effect is clear in both directions; the similar targets have a lower cross-entropy at each of the 20 ticks. Furthermore, priming is approximately equal in both directions.

Figure 5 (lower panel) presents the mean number of ticks required to reach each level of cross-entropy. A two-way repeated measures ANOVA that was identical to that of Simulation 1 was conducted. Prime-target similarity influenced convergence latency, $F(1, 37) = 27.45$. Convergence latency increased across the nine error levels, $F(8, 296) = 146.25$. Although similarity and error level again interacted, $F(8, 296) = 16.02$, planned



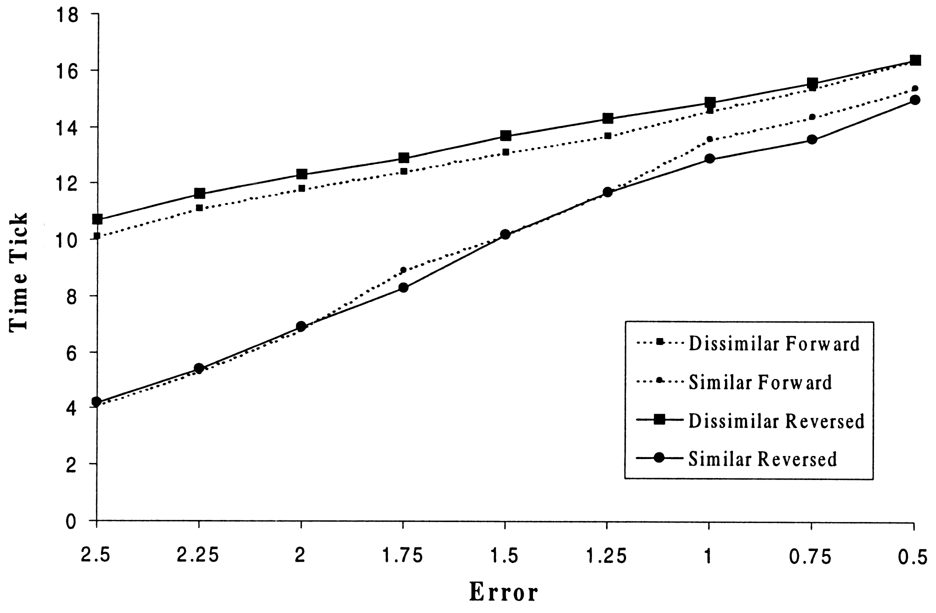


Figure 5. Upper panel: The settling profiles averaged over the 38 similar and dissimilar targets from MB Experiment 1 in both forward and reversed directions. Lower panel: The number of ticks required to reach various levels of cross-entropy.

comparisons revealed significant priming effects at all nine error levels. These results differed from the forward direction only in that the priming effect was nonsignificant at error level 0.5 in Simulation 1. Furthermore, paired-groups *t*-tests comparing the priming effects for Simulations 1 and 3 showed that they did not significantly differ at any of the nine error levels.

Discussion. The forward and reversed priming effects differed minimally from those of Simulation 1. Note, however, that they were not identical; priming in the reverse direction was slightly larger than in the forward direction. In contrast, if predictions were derived solely from similarity in terms of individual features, they would be identical in the two directions. These differences highlight the fact that the network's performance is influenced by factors such as the patterns of distinctive features in the two directions and the number of features and degree of structure among them in each concept.

Experiment 1

The purpose of Experiment 1 was to test whether the semantic-similarity priming effects found in MB Experiment 1 would be equivalent when prime-target presentation order was reversed, as predicted by the attractor network.

Method

Subjects. Thirty-six native English speaking psychology students at the University of Western Ontario participated for course credit; there were 18 in each list.

Materials. The 38 prime-target pairs were identical to MB Experiment 1, except that the primes served as targets, and vise versa. Two lists were created so that subjects saw no prime or target twice. For each list, 19 targets were paired with similar primes, and 19 with dissimilar primes. Dissimilar primes were created by re-pairing the similar primes and targets. The items are listed in Appendix A. All filler and practice items were identical to those used in the paired-presentation semantic decision task of MB Experiment 1.

In a norming study reported in MB Experiment 1, 28 subjects judged the similarity of the prime-target pairs, 14 in each direction. A paired-groups *t*-test with items as the random variable showed a small difference in the similarity ratings between the original ($M = 6.1$, $SE = .1$) and reversed ($M = 6.2$, $SE = .1$) directions, $t(37) = 1.05$; $p > .3$. Furthermore, the item-wise correlation for the ratings in the two directions was .84; $r^2 = .71$; $t(36) = 9.46$. Independent-groups *t*-tests with direction as the independent variable compared the targets in terms of length in letters, Kucera and Francis (1967) word frequency, and concept familiarity, as established by having 20 subjects rate "How familiar are the things that these words refer to?" on a 7-point scale where 7 was extremely familiar (originally reported in McRae et al., 1997). The reversed targets ($M = 5.3$, $SE = .3$) contained marginally fewer letters than did those of MB Experiment 1, ($M = 6.1$, $SE = .3$); $t(74) = 1.80$; $p < .08$. The reversed targets ($M = 16$, $SE = 4$) were also nonsignificantly more frequent than in the original direction, ($M = 11$, $SE = 3$); $t(74) = 1.06$;

$p > .2$. Rated concept familiarity was virtually identical for the reversed ($M = 5.0$, $SE = .2$) and original targets, ($M = 5.0$, $SE = .2$); $t(74) = 0.72$; $p > .4$. In summary, differential priming effects would not be expected on the basis of these stimulus characteristics.

Procedure. Subjects were tested individually using PsyScope (Cohen, MacWhinney, Flatt, & Provost, 1993) on a Macintosh LC630 with a 14-inch color Sony Trinitron monitor. They responded by pressing one of two buttons on a CMU button box. The subjects' index finger of their dominant hand was used for a "yes" response. A trial consisted of a fixation point "+" for 250 ms, followed by the prime for 200 ms, a mask (&&&&&&&&) for 50 ms, and then the target, which remained on screen until the subject responded. The ITI was 1500 ms. Subjects were instructed to read "the first word" and respond "only to the second word." Their task was to judge whether or not the second word referred to a concrete object. It took approximately 20 min to complete the task. All aspects of the procedure were identical to the paired-presentation semantic decision task of MB Experiment 1.

Design. The independent variable was similarity (similar versus dissimilar) and was within subjects (F_1) and items (F_2). A list factor (or item rotation group) was included as a dummy variable to stabilize variance caused by the rotation of items and subjects over the two lists (Pollatsek & Well, 1995). The dependent measures were decision latency and accuracy. To compare priming in the two directions, *t*-tests were conducted using priming effect as the dependent variable and experiment (MB Experiment 1 versus Experiment 1) as the independent variable. Experiment was between subjects and within items. Note that *t*-tests were used because it was felt this was more sensitive than testing for an interaction between relatedness and experiment.

Results

Decision Latencies. In the decision latency analyses reported in this article, trials on which an error occurred were excluded. Latencies greater than 3 *SD* above the grand mean were replaced by the cutoff value, 2% of the scores.

Decision latencies for semantically similar items ($M = 693$ ms, $SE = 20$ ms) were 45 ms shorter than for dissimilar items ($M = 738$ ms, $SE = 22$ ms); $F_1(1, 34) = 19.78$; $F_2(1, 36) = 28.51$. These results approximate those of the paired-presentation semantic decision task of MB Experiment 1 in which there was a 41 ms priming effect (similar: $M = 699$ ms, $SE = 22$ ms; dissimilar: $M = 740$ ms, $SE = 24$ ms); $t_1(76) = 0.25$; $p > .8$; $t_2(37) = 0.16$; $p > .8$, respectively.

Errors. Analyses were conducted using the square root of the number of errors as the dependent variable (Myers, 1979). Subjects made errors on 1.5% ($SE = .5\%$) of the similar prime-target trials and 1.3% ($SE = .4\%$) of the dissimilar trials, $F < 1$ in both analyses. This approximated MB Experiment 1 in which subjects made errors on 1.8% ($SE = .6\%$) of the similar and 2.1% ($SE = .6\%$) of the dissimilar trials, $t_1(76) = 0.88$; $p > .3$; $t_2(37) = 0.70$; $p > .4$, respectively.

Discussion

From Simulation 3, it was predicted that priming effects would change minimally when semantically similar non-associated prime-target pairs were presented in reverse order to subjects. This prediction was borne out in the human data of Experiment 1. Note that if factors such as word length, word frequency, and concept familiarity differed substantially in the two directions, differential priming effects could be obtained.

Experiment 1 does not differentiate between the predictions of an attractor network model of semantic memory and a spreading activation model. In a spreading activation network, such as that of Collins and Loftus (1975), semantic-similarity priming can occur via two mechanisms. They stated that featural similarity is one important dimension along which semantic memory is organized, and they implemented this notion by linking concepts through their shared features. Thus, priming can occur on the basis of the activation that spreads from one concept node to another through nodes representing their shared features. Although some links are directional in a semantic network, there seems to be no reason to think that these networks would predict directionally asymmetric priming effects for semantically similar, non-associated lexical concepts. In this way, the predictions of a connectionist attractor network and a spreading-activation network do not appear to differ.

The second way in which semantic-similarity priming can occur in some instantiations of semantic-network theory arises from hierarchical structure. For example, Collins and Quillian (1969) strongly emphasized the role of “isa” links in semantic processing (e.g., a *robin* isa *bird*). In their extension of this theory, they claimed that “Superordinate links act like highly criterial property links” (pp. 413–414), meaning that activation spreads strongly and quickly between exemplar and superordinate category nodes. Thus, priming is predicted to occur via activation spreading from the prime to the target through a shared superordinate category node (e.g., *truck* ⇒ *vehicle* ⇒ *van*). Again, however, priming should be equivalent in both directions.

VI. EXPERIMENT 2: FEATURAL OVERLAP VERSUS SHARED CATEGORY NODE

As just described, semantic-similarity priming can occur in spreading-activation networks, such as those of Collins and Loftus (1975), via links between shared features, shared superordinate category nodes, or a combination of the two. In contrast, because attractor networks do not contain superordinate category nodes, this type of priming must result from overlap in terms of individual features or the correlations among them. The purpose of Experiment 2 was to contrast these two notions of semantic similarity.

The majority of semantic-similarity priming studies have been based on assumptions derived from spreading-activation theory. Thus, lexical concepts, such as *truck* and *van*, are deemed similar if they are exemplars of the same superordinate category, such as *vehicle* (e.g., Chiarello et al., 1990; Hines et al., 1986; Lupker, 1984; Moss et al., 1995). On the other hand, three studies defined semantic similarity in terms of featural overlap

(MB; McRae et al., 1997; Shelton & Martin, 1992). Grouping previous experiments by this factor, however, does not provide insight into which view of semantic similarity best matches human performance. The first problem is that only two studies have found clear priming effects when items were not normatively associated and when subjects' strategies were minimized, and these studies (Chiarello et al., 1990; MB) differ in their definition of semantic similarity. Furthermore, although Chiarello et al. (1990) defined their items in terms of shared superordinate category, they presumably were highly featurally similar as well. Likewise, whereas MB defined their items in terms of semantic overlap, they were clearly category co-ordinates. Thus, the two variables have been confounded so that an explanation based on either mechanism is valid.

One way to determine whether shared superordinate category is the key factor is to test whether prime and target typicality influence priming. The degree to which an exemplar is typical of a category is a strong variable in many tasks (Erreich & Valian, 1979; Rosch & Mervis, 1975). Typicality is often considered as reflecting the similarity between an exemplar (*robin*) and its corresponding superordinate category concept (*bird*). Thus, the strength of the exemplar ⇔ superordinate link in a semantic network should be directly related to typicality, so that an account in which priming is mediated by a superordinate node predicts that priming between category co-ordinates depends on their typicality. Chiarello and Richards (1992) investigated this notion by comparing priming effects for typical (*robin-crow*) versus less typical (*duck-crow*) primes, with the two groups of items equated for rated featural similarity. They used an SOA of 575 ms, but no evidence of subjects' strategies was found. The results were mixed; equivalent priming effects for highly and less typical primes were found in lexical decision when primes and targets were presented in the left visual field, whereas no priming was found when words were presented to the right. In a pronunciation task, numerically but not significantly larger priming effects were found for the highly typical primes in both visual fields. Thus, it is not clear what to conclude from their experiments.

On the other hand, strong support for the centrality of featural similarity was reported by McRae et al. (1997). Their experiment included 54 artifact (*mittens-gloves*) and 34 living thing (*horse-pony*) prime-target pairs. The primes and targets were clearly category co-ordinates. On the basis of their feature norms, McRae et al. (1997) constructed semantic representations, in terms of individual features and correlated feature pairs, for these lexical concepts. Prime-target similarity was then computed in terms of shared and distinct individual and correlated features. Priming was measured by using a 250-ms SOA paired-presentation procedure with a semantic decision task. Regression analyses were conducted to predict item-by-item priming effects. Artifact priming effects were predicted by similarity in terms of individual features, whereas living thing priming effects were predicted by similarity in terms of correlated feature pairs (because features tend to be more densely intercorrelated for living things). Critically, featural similarity predicted the magnitude of priming effects *within* a category.

Further analyses were conducted on McRae et al.'s (1997) data to test for an influence of typicality (they collected typicality ratings for their items by having 20 subjects rate it on a 7-point scale for which 7 corresponded to extremely typical). Item-by-item priming effects were predicted by using prime typicality, target typicality, summed typicality, as well as similarity in terms of individual and correlated features. Two items were deleted because the typicality ratings for *shed-barn* and *crayon-pencil* were collected with respect to the superordinate *tool*, and the ratings showed that subjects did not consider them part of this category. When all 86 prime-target pairs were included in the regression analyses, similarity in terms of correlated features was the strongest predictor, $r^2 = .16$; $F(1, 83) = 15.60$, and similarity in terms of individual features also significantly predicted priming effects, $r^2 = .15$, $F(1, 83) = 14.10$. In contrast, none of the typicality measures predicted priming: prime typicality, $r^2 = .03$, $F(1, 83) = 2.37$, $p > .1$; target typicality, $r^2 = .01$, $F(1, 83) = 1.10$, $p > .2$; summed typicality, $r^2 = .03$, $F(1, 83) = 2.15$, $p > .1$. Note that because the variation in the typicality ratings was slightly greater than in the two similarity measures, the differences in predictive ability cannot be attributed to this factor.

One possibility for the lack of typicality effects is that superordinate categories for living things are more coherent than for artifacts (Atrans, 1989). In fact, Moss et al. (1995) hypothesized that this may have been the source of the stronger priming effects for living things than for artifacts in their study. Therefore, regression analyses were conducted separately for the 34 living thing and 52 artifact pairs. For the living things, as in McRae et al. (1997), similarity in terms of correlated features predicted priming effects, $r^2 = .23$; $F(1, 31) = 9.02$, whereas individual features did not, $r^2 = .04$; $F(1, 31) = 1.30$; $p > .3$. Again, none of the typicality measures predicted priming, $r^2 < .01$; $F < 1$ for all three measures. For the artifacts, similarity in terms of individual features predicted priming effects, $r^2 = .19$; $F(1, 49) = 11.27$, as did correlated features, $r^2 = .14$; $F(1, 49) = 7.79$. Note that this differs from McRae et al. (1997) because in their regression analyses, similarity in terms of individual features was always forced into the equation, and similarity in terms of correlated feature pairs was used to predict the residual variation. Again, none of the typicality measures predicted priming: prime typicality, $r^2 = .04$, $F(1, 49) = 1.90$, $p > .1$; target typicality, $r^2 = .03$, $F(1, 49) = 1.41$, $p > .2$; summed typicality, $r^2 = .04$, $F(1, 49) = 1.91$, $p > .1$. Finally, when the most highly correlated similarity measure was forced into the regression equation, the typicality variables did not predict the residual variation for the 86 items, the living things, nor the artifacts, $F < 1$ in all cases.

To provide further support for the featural similarity view, Experiment 2 was designed to contrast directly the notions of feature overlap and shared superordinate category. As described in Simulation 2, MB Experiment 3 obtained short SOA priming for highly similar but not less similar items. However, no norming was conducted to demonstrate that the members of the less similar priming pairs were category co-ordinates, and a casual inspection shows that they may have been for some of the pairs, but definitely not for all of them. Experiment 2 improved on MB Experiment 3 in that targets were paired with more similar/less typical and less similar/more typical primes (e.g., *squash* as the target, *pumpkin* as the more similar/less typical prime, *corn* as the less similar/more typical prime). Extensive norming was carried out to establish that subjects considered the members of each triplet to belong to the same category. In addition, the norms showed the less similar prime to be a better member of the category than its more similar counterpart.

If featural similarity is the key predictor of short SOA priming effects as predicted by an attractor network, then priming should obtain for only the more similar/less typical prime-target pairs. If priming occurs through a superordinate node, then priming may occur for both types of items, but it should be stronger for the less similar/more typical primes. A third possibility is predicted by Collins and Loftus (1975); priming should be relatively equal for both types of items because it is the product of both mechanisms. In this case, featural similarity plays the stronger role for the more similar/less typical items, whereas shared superordinate node is the source of priming for the less similar/more typical items.

Norming

Four norming studies produced 18 triplets (from 75 candidate triplets) that included a target, a more similar/less typical prime, and a less similar/more typical prime. No prime-target pairs were normatively associated. The 18 triplets are shown in Appendix B. Note that to construct these items, it was necessary to use primes and targets not included in the 190 normed concepts from the model. Therefore, although the attractor network clearly makes the prediction tested in Experiment 2, we could not directly simulate these data. In essence, Simulation 2 can be taken as the model's prediction.

Word Association Norms. To ensure that the prime-target pairs were not normatively associated, the experimenter read aloud either the targets or one set of primes from the initial 75 triplets. There were 48 subjects, 16 in each list. All subjects in the four norming studies were native English speaking University of Western Ontario undergraduates who received either course credit or cash remuneration for participating. Three lists were constructed; one contained all 75 target words, another the primes that the experimenters believed to be more similar and less typical, and a third contained the less similar/more typical primes. Three additional lists were constructed by reversing the order of the stimuli in each list. Subjects were asked to respond to each stimulus by saying the first word that came to mind. The experimenter read each word aloud and recorded the subject's responses. For the primes, we counted the number of times that the target item was produced. For the targets, we counted the number of times that either prime was produced. A word triplet was discarded if greater than one out of 16 subjects produced the target as a response to either prime, or vice versa, leaving 51 non-associated triplets.

Category Production Norms. Of the 47 subjects who participated in this study, two produced word associates rather than superordinate categories to many of the items and were discarded, leaving 15 subjects per word. Again, there were six lists; a forward and reverse version of lists containing either the targets or one prime type. The experimenter read instructions to each subject, and examples were provided. The experimenter then read each item aloud, and the subject indicated the category to which he or she believed the concept belonged. "I don't know" was considered a valid response. The most frequent response was designated as the item's dominant category. Eighteen triplets were retained for which the dominant category was identical for the target and less similar/more typical

TABLE 2
Characteristics of Experiment 2 Stimuli

		Prime	
Stimulus Characteristic	Target	More similar/less typical	Less similar/more typical
Similarity to target		6.0 (0.3)	4.4 (0.3)
% Dominant category responses	72.3 (3.5)	70.1 (3.3)	83.0 (2.4)
Typicality	6.6 (0.3)	6.8 (0.3)	7.8 (0.2)
Length in letters	6.5 (0.4)	5.4 (0.3)	5.3 (0.3)
Word frequency*	7.7 (2.7)	10.2 (3.7)	16.3 (6.5)

Note. SE in parentheses. *According to Kucera and Francis (1967).

prime and at least 50% of the subjects produced it. The mean percentage of dominant category responses are presented in Table 2. ANOVAs were conducted with item type as the independent variable (more similar/less typical prime versus less similar/more typical prime versus target) and percentage of dominant category responses as the dependent variable. Item type was between subjects and within items. There was a main effect of item type, $F_1(2, 42) = 4.92$; $F_2(2, 34) = 7.99$. Planned comparisons revealed that the mean percentage of dominant category responses was greater for the less similar/more typical primes than for the more similar/less typical ones, $F_1(1, 42) = 8.59$; $F_2(1, 34) = 13.85$.

For a few of the 18 items, a secondary superordinate category name was produced by more than one subject. In terms of predicting priming effects, a concern arises if the secondary category is smaller (less inclusive) than the dominant category because it could then be the case that priming might be mediated by that closer superordinate node. This situation arose for the *waffle-toast-pancake* triplet only, where *food* was the dominant category and *breakfast food* was the less inclusive secondary category. Subjects produced *breakfast food* 33% of the time to *waffle* and *pancake*, and 7% of the time to *toast*, which was the less similar/more typical prime.

Typicality Ratings. A separate set of 17 subjects rated the typicality of the 54 exemplars (18 triplets) in a paper-and-pencil format. Each exemplar was included with its dominant category, as determined by the previous norming study. The category-exemplar pairs were presented in sentences of the form: “How typical of a VEGETABLE is CORN?” Beside each sentence appeared a 9-point scale, with a score of 1 corresponding to not at all typical and a score of 9 corresponding to extremely typical. Each subject was given a copy of the rating scale, and the experimenter read aloud standard typicality rating instructions. An example was provided, and any questions were answered. Subjects completed the task without a time constraint.

Mean typicality ratings appear in Table 2. ANOVAs were conducted with item type as the independent variable (more similar/less typical prime versus less similar/more typical prime versus target) and typicality rating as the dependent variable. Item type was within subjects and items. There was a main effect of item type, $F_1(2, 32) = 41.41$; $F_2(2, 34) = 4.55$. Planned comparisons showed that the mean typicality rating was greater for the less similar/more typical primes than for the more similar/less typical ones, $F_1(1, 32) = 45.83$; $F_2(1, 34) = 4.95$.

Similarity Ratings. In the final norming study, 36 subjects were shown either the more similar/less typical prime-target pairs or the less similar/more typical prime-target pairs combined with those of Moss et al. (1995) and Shelton and Martin (1992). The study was conducted in this manner to avoid repeating concepts, to have the same filler concepts for each set of items because ratings are sensitive to list context, and to allow direct comparison to Moss et al. (1995) and Shelton and Martin (1992).

Items were presented in a paper-and-pencil format in sentences of the form, "How similar are SQUASH and PUMPKIN?" or "How similar are PUMPKIN and SQUASH?" Nine subjects were presented with the prime-target pairs in each order. Beside each sentence appeared a 9-point scale in which 1 corresponded to not at all similar and 9 to extremely similar. Subjects were instructed to rate the similarity of the things to which the words referred. There was no time constraint.

Mean similarity ratings also appear in Table 2. Because the two types of items were presented in different lists and the Shelton and Martin (1992) and Moss et al. (1995) items were in both lists, separate analyses were conducted. The more similar/less typical pairs were more similar than those of Moss et al. (1995), ($M = 4.6$, $SE = .2$), $F_1(1, 16) = 161.02$; $F_2(1, 77) = 17.16$, and those of Shelton and Martin (1992), ($M = 3.8$, $SE = .2$), $F_1(1, 16) = 387.30$; $F_2(1, 77) = 45.42$. The less similar/more typical items were judged to be as similar as those of Moss et al. (1995) ($M = 4.5$, $SE = .2$), $F_1(1, 16) = 1.50$; $p > .2$; $F_2 < 1$, but more similar than those of Shelton and Martin (1992), ($M = 3.7$, $SE = .2$), $F_1(1, 16) = 35.00$; $F_2(1, 77) = 5.26$. Finally, we compared the Experiment 2 items directly by omitting all others. The more similar/less typical prime-target pairs were indeed more similar, $t_1(34) = 4.79$; $t_2(17) = 6.43$.

Short SOA Priming

The 18 triplets were used in a short SOA paired-presentation lexical decision task that was designed to determine whether semantic-similarity priming is mediated by a shared category node, is a function of featural similarity, or both.

Method

Subjects. Sixty-six University of Western Ontario undergraduates participated for either course credit or a cash reward. All participants were native speakers of English and had normal or corrected-to-normal vision. Twenty-two subjects were randomly assigned to each of the three lists.

Materials. Three lists were created so that subjects saw no prime or target twice. For each list, six targets were paired with more similar/less typical primes, six with less similar/more typical primes, and six with unrelated primes. Unrelated primes were selected by re-pairing high similarity primes and targets. There were 102 filler trials

TABLE 3
Mean Lexical Decision Latency in ms and Percent Errors for Experiment 2

Prime type	Decision latency	Percent errors
More similar/less typical	636 (12)	6.2 (1.1)
Less similar/more typical	658 (12)	7.3 (1.2)
Unrelated	662 (12)	9.7 (1.2)

Note. SE in parentheses.

consisting of 42 unrelated word-word and 60 word-nonword pairs per list. The relatedness proportion was .2, and the nonword ratio was .56.

Procedure. All aspects of the procedure were identical to Experiment 1 except that a lexical decision task was used because there was not one semantic question that encompassed all of the items, and there were not enough items to break the task down into multiple semantic decisions. Subjects were given 40 practice trials before the 120 experimental trials, which were presented in random order. Two rest periods were provided during the experimental trials.

Design. The independent variable was prime type (more similar/less typical versus less similar/more typical versus unrelated). A list factor (or item rotation group) was again included. Prime type was within subjects and items. The dependent measures were decision latency and accuracy.

Results

Mean decision latency and error rate for each condition are presented in Table 3. Latencies greater than 3 *SD* above the grand mean were replaced by the cutoff value, 1% of the scores.

Decision Latencies. Lexical decision latencies differed by prime type, $F_1(2, 162) = 4.26$; $F_2(2, 30) = 4.75$. Planned comparisons revealed that subjects responded 26 ms faster to the more similar/less typical prime-target pairs than to the unrelated pairs, $F_1(1, 162) = 7.34$; $F_2(1, 30) = 8.72$. Furthermore, subjects responded 22 ms faster to the more similar/less typical pairs than to the less similar/more typical pairs, $F_1(1, 162) = 5.29$; $F_2(1, 30) = 5.02$. The 4-ms priming effect for the less similar/more typical pairs was not reliable, $F_1 < 1$, $F_2 < 1$.

Error Rates. No differences were significant.

Discussion

Mean decision latency was shorter for targets preceded by more similar/less typical primes than for those preceded by either less similar/more typical primes or unrelated primes. Furthermore, there was a small, non-significant difference between the latter two conditions, demonstrating that featural overlap rather than shared superordinate category accounts for semantic-similarity priming effects. These results refute a central aspect of spreading-activation theories that propose hierarchical levels of representation; the results strongly suggest that semantic memory does not consist of a set of concept nodes organized in a hierarchical fashion. The hierarchical nature of semantic memory and the key role played by superordinate nodes are critical components of semantic-network theories, such as those of Collins and Quillian (1969) and Collins and Loftus (1975). Of course, this experiment does not completely discount spreading-activation models because semantic-similarity priming effects can be attributed to featural links between highly similar concept nodes.

In addition to empirical problems with an account of semantic memory that emphasizes hierarchical semantic structure that is coded in terms of local category nodes, there are logical problems as well. For instance, there are inherent difficulties in determining what superordinates are psychologically real, and hence, what category nodes would be implicated to play a role in semantic priming. Many types of concepts exist for which the relevant superordinates are not obvious, particularly to the average person. These might include verbs, such as *run* or *break*, adjectives, such as *silent* or *beautiful*, and even concrete nouns, such as *fence* or *garage*. This problem was apparent when conducting the category production norming task because numerous basic-level concepts did not induce consistent superordinate category responses. Barsalou (1987) argued, along similar lines, that superordinate categories should not be viewed as static nodes in a hierarchically organized semantic system. Rather, on the basis of the variation in typicality ratings across individuals and within individuals over time, as well as the results showing that people treat ad hoc categories, such as *things on my desk*, in much the same way as taxonomically-based categories, he concluded that people's representations of categories are not stable entities, but are computed only when needed and constantly change as a result of experience. In summary, semantic similarity cannot be defined in terms of a shared category node. Rather, it results from the overlap of individual features and higher-order semantic structure that is a natural consequence of a distributed representation of word meaning.

Although attractor networks do not contain superordinate nodes, this does not imply that higher-level categories are irrelevant to their functioning. In fact, the ways in which concepts form coherent clusters are important to the dynamics of these models. For example, most sets of feature intercorrelations occur within superordinate categories, such as *vegetable* or *weapon*, and these semantic regularities almost certainly play a role in learning the meaning of these words. Although our model was not trained to learn the meaning of such superordinate terms, there are potential ways of achieving this that we are currently exploring. One possible method is to have a network learn to map from a superordinate category label, such as *vegetable*, to the semantic representations of the basic-level objects that are in that category (e.g., *carrot*, *corn*, *lettuce*). This type of learning trial would be interspersed with trials in which the basic-level label and semantic representation would be presented together. A major challenge in implementing this approach is to determine the relative frequencies with which people refer to an object with the basic-level versus superordinate labels. These frequencies presumably are reflected in variables such as typicality ratings and the frequency with which subjects produce a superordinate label in a feature norming or category production task. In addition, when people speak, whether they use a superordinate or a basic-level label may be related to the availability of the lexical form, therefore, the less familiar a person is with the basic-level concept and label, the more likely he or she is to substitute the superordinate. When testing a network trained in this manner, one would expect, for example, that the computed representations would differ in terms of clarity (i.e., the number of features that are unambiguously on or off, rather than being ambiguous between these states). The clarity of the superordinate semantic representations should be influenced primarily by the shared individual and correlated features among the exemplars for which the label was substituted during training [for a discussion of how superordinate representations might fall out of Hinton & Shallice's (1991) featural representations, see Small, Hart, Nguyen, & Gordon, 1995]. Note that effects of clarity of representation have been found in feature production norms. For example, subjects have much less difficulty providing features for *fruit* than for *furniture* (McRae et al., 1997; Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976). In summary, although attractor networks do not contain superordinate category nodes, the semantic structure that superordinates reflect plays a role in their computational dynamics.

VII. GENERAL DISCUSSION

The modeling and experiments presented herein contribute to the literatures on semantic memory and semantic priming by testing an attractor network that instantiates a theory in which similarity defined over individual features and feature intercorrelations is a major organizing principle of word meaning. Simulation 1 showed that such a network can account for a basic semantic-similarity priming effect. Simulation 2 reproduced the more subtle effects of degree of similarity that have been used to elucidate the factors that control their magnitude. More important, empirically derived semantic representations were used in the model to reduce the degrees of freedom in the simulations. Two previously untested predictions were then derived from the model and tested with human subjects. First, although it had been predicted that semantic-similarity priming effects should be roughly equivalent regardless of prime-target order, it had not been tested with a model nor humans. The prediction held in the model in Simulation 3 and with humans in Experiment 1. Second, following the basic organizational principles of the network and the results of Simulation 2, it was predicted that semantic-similarity priming is due to featural overlap, rather than shared superordinate category node. This prediction was confirmed in Experiment 2, which conflicts with a central tenet of hierarchical spreading-activation networks, from which most previous hypotheses concerning semantic priming had been derived. In summary, attractor networks provide a natural account of the major behavioral phenomena associated with semantic-similarity priming.

Limitations, Extensions, and Other Issues

This article focused on semantic-similarity priming in tasks in which subjects' strategies are assumed to be minimized. This approach was taken because we believe that we have a clear understanding of the behavioral phenomena in this domain, how semantic similarity is naturally incorporated into a model, and how to simulate the relevant experiments. However, the semantic priming literature is expansive, and accounting for the complete range of behavioral phenomena requires a more complex model than the one presented herein. In the remainder of the article, we discuss a number of variables that were not incorporated into our model, outline how they could be, and point toward some new avenues of research that become apparent when implementing a model of this sort.

Semantic Relatedness. Numerous experiments have used lexical concepts that were construed broadly as being semantically related. General semantic relatedness was not incorporated into the model because this category encompasses numerous relationships, such as featural overlap, concept-feature (*dog-tail*), superordinate-exemplar (*vehicle-truck*), and script relations (*restaurant-wine*), to name a few. Thus, it is likely that multiple mechanisms will be required to implement various semantic relationships in a realistic fashion and that priming effects depend on the type of relation in both model simulations and human experiments (Moss et al., 1995). Therefore, to obtain a clearer picture of semantic relations, they must be identified, clarified, implemented in a model, and investigated in human experiments. That the vast majority of published experiments use a mixture of items that exemplify a number of types of relations can be converted to a strength for the modeling enterprise; simulations can generate predictions for new experiments because the relevant data does not exist. This article, in conjunction with McRae et al. (1997) and MB is an example of this type of research, focusing on featurally similar basic-level concepts.

A number of semantic relations could be naturally incorporated into a network such as ours. For example, concept-feature and feature-feature relations already exist in the network and could be tested if individual features were trained with corresponding word forms. Antonyms and superordinate-exemplar relations, as forms of semantic similarity, would also emerge from a model of this sort. However, it may not be so straightforward to simulate all types of semantic relations. For example, Moss et al. (1995) found robust priming for functional relations, such as *broom-floor*, and priming for script relations, such as *restaurant-wine,* was obtained in some of their conditions but not others. Another example of script- or event-based relations is evident in the research of Ferretti, McRae, and Hatherell (submitted for publication) who found that verbs prime their typical agents (*arresting-cop*), patients (*arresting-criminal*), and instruments (*stirred-spoon*). More significant extensions to the model are required to incorporate these relations and to simulate the relevant data.

Associative Relatedness and Co-occurrence. Unlike semantic relatedness, associative relatedness has been defined in a precise manner. Its definition consists of the operationalization of word association norms: if one word is produced in response to another in a word association task, then they are related. However, severe problems exist with this operationalization because of the numerous bases for subjects' responses in the word association task, including the semantic relations listed in the previous section (Hodgson, 1991; McKoon & Ratcliff, 1992). Recently, a number of researchers have suggested that this problem might be alleviated if associative relatedness was defined in terms of co-occurrence frequency in speech or text, as measured by analyses of language corpora (McKoon & Ratcliff, 1992; Williams, 1996). For example, Williams (1996) found priming in short SOA lexical decision and naming tasks for collocates, such as *bread* and *butter*, *salt* and *pepper*, and *aunt* and *uncle*. Note that if the collocates refer to objects, as did all of Williams' (1996) words, the objects probably occur together in the world (this, presumably, being the underlying reason for why they co-occur in language). One reason why the notion of co-occurrence is attractive is that it is readily modeled in general (for a discussion of incorporating co-occurrence into compound-cue theory, see McKoon & Ratcliff, 1992) and can easily be incorporated into an attractor network by modifying the training regime. In fact, Plaut (1995) incorporated this relation during training by having items follow one another with a reasonably high probability [note, however, that Plaut (1995) presented this manipulation as modeling associative relatedness in general]. Plaut (1995) found that the backpropagation-through-time learning algorithm is sensitive to temporal contiguity in that attractors develop so that the second item of a pair converges more quickly when it is preceded by its collocate than by an unrelated word.

Problems with defining and measuring association also bear on the present experiments; we used association norms to establish that the semantically similar items were not associatively related. Although this is the accepted method in the literature, there are questions about whether counting the number of primary associates and removing any items that exceed some conservative criterion is an adequate method for factoring out association. There are two possible responses. First, MB suggested that this method may be overly conservative. Because semantic similarity is one of the relations driving performance on the word association task, it is possible that removing items on the basis of this task simply results in excluding the best stimuli. In contrast, if a researcher views association norms as reflecting the strength of associative links in semantic memory, it might be argued that subjects should be asked to produce multiple associates so that weakly associated items can be factored out. However, this version of the word association task is severely flawed because it is highly likely that subjects associate not only to the stimulus, but also to their previous responses. Therefore, this method compounds the problems involved with using association norms for this purpose.

An alternative method for distinguishing between featural overlap and associative relatedness is to exclude items that co-occur in speech and text. On the basis of this logic, we tested whether the items of Experiments 1 and 2 co-occurred to a significant extent by using the text corpus on which the Hyperspace Analog to Language (HAL) model of Burgess, Lund, and colleagues is based (for a description of the basic HAL methodology, see Lund & Burgess, 1996). The statistics presented herein are based on a HAL matrix in which each cell contained information concerning the directional co-occurrences of two words within a 10-word window. For the purposes of removing co-occurring words from priming stimuli, a 10-word window probably provides a liberal estimate. The actual co-occurrence in HAL for two words, such as *eagle* and *hawk*, is weighted by their distance; it is incremented by 10 if they are adjacent, 9 if one word intervenes, and so on. Each co-occurrence was included in the matrix only once (i.e., it is not counted multiple times as the window is moved one word at a time). Given these co-occurrence statistics and the frequency of each word (out of approximately 300,000,000 words), we computed the relative occurrence of a target given a prime, as in Equation 4, where *freq* refers to the word's frequency.

$$\frac{freq(\text{target}|\text{prime})}{freq(\text{prime})} \qquad (4)$$

This statistic provides an indication of the probability that the target follows the prime in any 10-word window in which the prime occurs.

We computed this co-occurrence statistic for the items of Experiment 2. Because some of the words did not appear in the HAL database, statistics were calculated for 15 of the 18 word triplets. A one-way ANOVA with co-occurrence scores as the dependent variable and prime type (more similar/less typical versus less similar/more typical versus unrelated) as the within-items independent variable revealed a nonsignificant effect of prime type, $F_2(2, 28) = 2.48$; $p > .1$. Planned comparisons revealed that co-occurrence scores for the more similar/less typical items were greater than for the unrelated items, $F_2(1, 28) = 4.30$, and were marginally greater than for the less similar/more typical items, $F_2(1, 28) = 3.04$; $p > .09$, with no difference between the less similar/more typical and unrelated items, $F_2 < 1$. The differences between groups were caused by two outlying items: the co-occurrence scores for *peas-beans* and *nutmeg-cinnamon* were seven times greater than for the prime-target pair with the third highest score, *pumpkin-squash*. Thus, the co-occurrence scores and Experiment 2 decision latencies were re-analyzed after excluding these items. We also removed *pumpkin-squash* because doing so resulted in an equal number of items in each rotation group of Experiment 2. After eliminating the outliers, there were no differences in co-occurrence scores [more similar/less typical: $M = .002$, $SE = .001$; less similar/more typical: $M = .002$, $SE = .002$; unrelated: $M = .001$, $SE = .001$; all F_2s < 1]. Furthermore, the pattern of Experiment 2 decision latencies did not change; there was a main effect of prime type, $F_2(2, 24) = 4.81$, with the more similar/less typical items ($M = 648$ ms, $SE = 27$ ms) having shorter decision latencies than both the less similar/more typical items ($M = 679$ ms, $SE = 29$ ms), $F_2(1, 24) = 7.52$, and the unrelated items ($M = 680$ ms, $SE = 26$ ms), $F_2(1, 24) = 6.88$. The 1 ms difference between the latter two groups was not significant, $F_2 < 1$. In summary, although the co-occurrence statistic differed among groups because of the small number of items, the decision latency results did not change from Experiment 2 when those items were eliminated.

As a further illustration that semantic-similarity priming effects definitely do occur in the absence of association (co-occurrence), we conducted an ANOVA that used only the

10 items from Experiment 1 for which the prime *never* preceded the target in a 10-word window. Semantic decision latencies for the similar prime-target pairs ($M = 699$ ms, $SE = 37$ ms) were 74 ms shorter than for the dissimilar pairs ($M = 773$ ms, $SE = 36$ ms), $F_2(1, 8) = 45.38$. This test is extremely conservative because co-occurrence frequencies as measured in a 300,000,000 word corpus would presumably need to be substantial (i.e., much greater than a single or a few co-occurrences) for this factor to influence lexical processing sufficiently to result in facilitation in a priming task. In fact, an interesting empirical question concerns whether it is possible to obtain priming for items that co-occur frequently, but share no other obvious semantic relationship. The experiment that has come closest to answering this question is McKoon and Ratcliff (1992, Experiment 3), who obtained priming in a single-presentation lexical decision task by using words that co-occurred frequently in a text corpus. The vast majority of their items were phrasal associates, such as *public-health* and *movie-stars*.

Other Variables. A number of other variables have been manipulated in priming experiments and should be taken into account when simulating behavioral phenomena. For example, the type of task subjects perform on the target has been shown to influence priming effects, with most reported studies using either lexical decision or naming. In studies that have used the same items in lexical decision and naming tasks, smaller effects are typically found with naming, with this difference being attributed to the strategic effects that are hypothesized to influence lexical decisions (Seidenberg et al, 1984; Neely, 1991). Simulating all three tasks that have been used in priming experiments, namely semantic decision, lexical decision, and naming, requires a model that includes orthographic, phonological, and semantic representations, as in Masson (1995), whereas our model included word meaning and an abstract representation of word form.

There appear to be a number of factors that vary among tasks that are relevant to the expected magnitude of priming effects. The one that has received the most attention is retrospective effects that are argued to influence lexical decisions but not naming (Neely, 1991). Because it has been argued that these effects influence the decision process, rather than the lexical computations themselves, these might be modeled by supplementing a network with a binary decision process. The second factor is the match between the degree to which performance on a task is influenced by specific lexical computations and the degree to which those computations are influenced by the relation(s) exemplified in the materials. For example, in an attractor network, co-occurrence influences the computation of all aspects of lexical representation so that priming should be found in semantic decision, lexical decision, and naming. Although the data are not clear on this point, studies such as Seidenberg et al. (1984) suggest that this is so. In contrast, semantic similarity is encoded in meaning only, not in orthography or phonology, so little or no priming of this sort would be expected to occur in naming. In concert with this hypothesis, in an unpublished experiment in our laboratory, no evidence for priming in a short SOA naming task using semantically similar non-associated items that showed robust priming in a semantic decision task was found. This prediction requires qualification, however, in that studies such as Strain et al. (1995) showed that semantic factors can influence naming if the task is made more difficult. That is, if naming latencies are long, the computation of word meaning has an opportunity to influence the computation of phonology. Williams (1996) supported this notion by obtaining short SOA priming for featurally similar items (that he termed "coordinates") in a naming task only when naming latencies were slowed by severely degrading the targets.

Finally, there are a number of other priming effects documented in the literature that probably require assumptions to be added to present connectionist models. These deal with effects that are typically attributed to subjects' strategies, such as those of non-word proportion and relatedness proportion found with long SOAs. As stated earlier, it is not clear how to implement task-specific strategies associated with long SOAs in a model such as ours. We consider our network to be an instantiation of a theory of how word meaning is computed, not a model of priming per se. It is reasonable to assume that implementing a model of a specific laboratory task should require additional assumptions. This is certainly the case for a variable such as non-word proportion that is relevant only to priming effects in lexical decision.

In summary, an important issue is whether a single attractor network of lexical processing could be constructed that would account for all of the major results in this domain. At present, no one model has been implemented and used to do so. However, aggregating over a few recent attractor networks that have been used to simulate priming phenomena suggests that this can be accomplished (Becker et al., 1997; Masson, 1995; McRae et al., 1997; Plaut, 1995). Perhaps more important, however, attractor networks may point to new ways to think about variables such as prime-target relations, and clearer predictions regarding factors such as the type of task performed on the target.

VIII. CONCLUSION

In conclusion, the research described herein provides further evidence for an attractor network theory of lexical processing. Attractor networks have now been used to account for a number of aspects of orthographic, phonological, and semantic processing in both normal and neurologically impaired individuals (e.g., Farah & McClelland, 1991; Kawamoto, 1993; Plaut et al., 1996; Plaut & Shallice, 1993). They have led to new insights and predictions that have been confirmed by subsequent experimentation (e.g., Devlin et al., 1998; McRae et al., 1997). Further development, integration, and testing of attractor networks will no doubt lead to a more detailed understanding of the representations and computations involved in language comprehension and production.

Acknowledgments: We thank Chad Audet, Curt Burgess, and Kevin Lund for providing us with the HAL statistics required to derive the measures used in this article. This work was supported by Natural Sciences and Engineering Research Council (Canada) grant OGP0155704 to the second author, and an Ontario Graduate Scholarship postgraduate fellowship to the first author. Part of this research formed part of GSC's Master's thesis. We thank Mike Masson and Dave Plaut for comments on an earlier version of this article.

NOTES

1. A dummy trial that was not analyzed was inserted at the beginning of each set of priming trials so that when the first prime was presented, the network was in the state corresponding to the dummy trial target.
2. The precise correspondence between time ticks in the network and a specific SOA is not clear. All simulation results reported herein used 15 ticks of the prime, but the results did not differ appreciably from about 11 or 12 ticks up to about 20 ticks. From observing the settling profiles of the primes, our intuition is that 15 prime ticks is probably a reasonable estimate of a 250-ms SOA, although the precise correspondence would depend on a subject's reading ability and other factors, such as prime familiarity and length, that are not included in the current version of the network.
3. In all analyses reported in this article, $p < .05$ unless otherwise noted.

REFERENCES

Anderson, J. R. (1983). *The architecture of cognition*. Cambridge, MA: Harvard University Press.

Atrans, S. (1989). Basic conceptual domains. *Mind and Language, 4,* 7–16.

Barrett, S. E., Abdi, H., Murphy, G. L., & Gallagher, J. M. (1993). Theory-based correlations and their role in children's concepts. *Child Development, 64,* 1595–1616.

Barsalou, L. W. (1987). The instability of graded structure in concepts. In Neisser, U. (Ed.), *Concepts and conceptual development: Ecological and intellectual factors in categorization* (pp. 101–140). Cambridge, MA: Cambridge University Press.

Becker, C. A. (1979). Semantic context and word frequency effects in visual word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 5,* 252–259.

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

Becker, S., Moscovitch, M., Behrmann, M., & Joordens, S. (1997). Long-term semantic priming: A computational account and empirical evidence. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 1059–1082.

Chiarello, C., & Richards, L. (1992). Another look at categorical priming in the cerebral hemispheres. *Neuropsychologia, 30,* 381–392.

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 and Language, 38,* 75–104.

Cohen, J. D., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: A new graphic interactive environment for designing psychology experiments. *Behavioral Research Methods, Instruments, & Computers, 25*, 257–271.

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

Collins, A. M., & Quillian, M. R. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior, 8,* 240–247.

Cottrell, G. W., Munro, P., & Zipser, D. (1987). Learning internal representations from gray-scale images: An example of extensional programming. *Proceedings of the Ninth Annual Conference of the Cognitive Science Society, 9,* 461–473.

de Groot, A. M. B. (1984). Primed lexical decision: Combined effects of the proportion of similar prime-target pairs and the stimulus-onset asynchrony of prime and target. *The Quarterly Journal of Experimental Psychology, 36A*, 253–280.

den Heyer, K., Briand, K., & Dannenbring, G. L. (1983). Strategic factors in a lexical-decision task: Evidence for automatic and attention-driven processes. *Memory & Cognition, 11,* 374–381.

Devlin, J. T., Gonnerman, L. M., Andersen, E. S., & Seidenberg, M. S. (1998). Category specific semantic deficits in focal and widespread brain damage: A computational account. *Journal of Cognitive Neuroscience, 10*, 77–94.

Dosher, B. A., & Rosedale, G. (1989). Integrated retrieval cues as a mechanism for priming in retrieval from memory. *Journal of Experimental Psychology: General, 118,* 191–211.

Erreich, A., & Valian, V. (1979). Children's internal organization of locative categories. *Child Development, 50,* 1071–1077.

Farah, M. J., & McClelland, J. L. (1991). A computational model of semantic memory impairment: Modality specificity and emergent category specificity. *Journal of Experimental Psychology: General, 120,* 339–357.

Fischler, I. (1977). Semantic facilitation without association in a lexical decision task. *Memory & Cognition, 5,* 335–339.

Forster, K. I. (1994). Computational modeling and elementary process analysis in visual word recognition. *Journal of Experiment Psychology: Human Perception and Performance, 20,* 1292–1310.

Gelman, S. A. (1988). The development of induction within natural kind and artifact categories. *Cognitive Psychology, 20,* 65–95.

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

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 and Performance, 12,* 370–379.

Hinton, G. E. (1989). Connectionist Learning Procedures. *Artificial Intelligence, 40,* 185–234.

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

Hodgson, J. M. (1991). Informational constraints on pre-lexical priming. *Language and Cognitive Processes, 6,* 169–264.

Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Science, 79,* 2254–2558.

Hopfield, J. J. (1984). Neurons with graded response have collective computational features like those of two-state neurons. *Proceedings of the National Academy of Science, 81,* 3088–3092.

Jones, S. S., & Smith, L. B. (1993). The place of perception in children's concepts. *Cognitive Development, 8,* 113–139.

Joordens, S, & Becker, S. (1997). The long and short of semantic priming effects in lexical decision. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 1083–1105.

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.

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.

Koriat, A. (1981). Semantic facilitation in lexical decision as a function of prime-target association. *Memory & Cognition, 9,* 587–598.

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

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

Lund, K., Burgess, C., & Atchley, R. A. (1995). Semantic and associative priming in high-dimensional semantic space. *Proceedings of the Seventeenth Annual Conference of the Cognitive Science Society, 17,* 660–665.

Lupker, S. J. (1984). Semantic priming without association: A second look. *Journal of Verbal Learning and Verbal Behavior, 23,* 709–733.

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

McNamara, T. P., & Altarriba, J. (1988). Depth of spreading activation revisited: Semantic mediated priming occurs in lexical decisions. *Journal of Memory and Language, 27,* 545–559.

McKoon, G., & Ratcliff, R. (1992). Spreading activation versus compound cue accounts of priming: Mediated priming revisited. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 1155–1172.

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

McRae, K., de Sa, V., & Seidenberg, M. S. (1997). On the nature and scope of featural representations of word meaning. *Journal of Experimental Psychology: General, 126,* 99–130.

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

Moss, H. E., Ostrin, R. K., Tyler, L. K., & Marslen-Wilson, W. D. (1995). Accessing different types of lexical semantic information: Evidence from priming. *Journal of Experimental Psychology: Learning, Memory and Cognition, 21,* 863–883.

Myers, J. L. (1979). *Fundamentals of experimental design*. Boston, MA: Allyn and Bacon.

Neely, J. H. (1977). Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited-capacity attention. *Journal of Experimental Psychology: General, 106,* 226–254.

Neely, J. H. (1991). Semantic priming effects in visual word recognition: A selective view of current findings and theories. In Besner D., & Humphreys, G. (Eds.), *Basic processes in reading: Visual word recognition* (pp. 264–336). Hillsdale, NJ: Erlbaum.

Peterson, R. R., & Simpson, G. B. (1989). The effect of backward priming on word recognition in single-word and sentence contexts. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 1020–1032.

Plaut, D. C. (1995). Semantic and associative priming in a distributed attractor network. *Proceedings of the Seventeenth Annual Conference of the Cognitive Science Society, 17,* 37–42.

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

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. E. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review, 103,* 56–115.

Pollatsek, A., & Well, A. D. (1995). On the use of counterbalanced designs in cognitive research: A suggestion for a better and more powerful analysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 785–794.

Posner, M. I., & Snyder, C. R. R. (1975). Facilitation and inhibition in the processing of signals. In Rabbitt, P. M. A., & Dornic, S. (Eds.), *Attention and performance V* (pp. 669–682). New York: Academic Press.

Postman, L., & Keppel, G. (1970). *Norms of word associations*. San Diego, CA: Academic Press.

Ratcliff, R., & McKoon, G. (1988). A retrieval theory of priming in memory. *Psychological Review, 95,* 385–408.

Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology, 7,* 573–605.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology, 8,* 382–439.

Rumelhart, D. E., Durbin, R., Golden, R., & Chauvin, Y. (1995). Backpropagation: The basic theory. In Rumelhart, D. E., & Chauvin, Y. (Eds.), *Backpropagation: Theory and practice* (pp. 1–34). Cambridge, MA: MIT Press.

Schwanenflugel, P. J., & Rey, M. (1986). Interlingual semantic facilitation: Evidence for a common representational system in the bilingual lexicon. *Journal of Memory and Language, 25,* 605–618.

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

Seidenberg, M. S, Waters, G. S., Sanders, M., & Langer, P. (1984). Pre- and post-lexical loci of contextual effects on word recognition. *Memory & Cognition, 12,* 315–328.

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

Small, S. L., Hart, J., Nguyen, T., & Gordon, B. (1995). Distributed representations of semantic knowledge in the brain. *Brain, 118,* 441–453.

Smith, E. E., Shoben, E. J., & Rips, L. J. (1974). Structure and process in semantic memory: A feature model for semantic decisions. *Psychological Review, 81,* 214–241.

Strain, E., Patterson, K., & Seidenberg, M. S. (1995). Semantic effects in single-word naming. *Journal of Experimental Psychology: Learning, Memory and Cognition, 21,* 1140–1154.

Tversky, A. (1977). Features of similarity. *Psychological Review, 84,* 327–352.

Williams, J. N. (1996). Is automatic priming semantic? *European Journal of Cognitive Psychology*, *8,* 113–161.

APPENDIX A
Prime-target pairs from Simulation 2 and Experiment 1 (reversed from MB Experiment 1)

Dissimilar prime	Similar prime	Target
barn	sandals	slippers
ship	camisole	bra
bomb	belt	tie
camisole	carpet	mat
sandpaper	pillow	cushion
pineapple	dresser	closet
beans	chandelier	lamp
budgie	jar	bottle
prune	toaster	microwave
pillow	sandpaper	file
subway	shovel	hoe
carpet	barn	shed
spear	pencil	crayon
canary	dunebuggy	jeep
chicken	scooter	motorcycle
caribou	van	truck
shovel	subway	bus
pencil	raft	canoe
sandals	ship	yacht
squash	cart	wagon
turkey	tomahawk	axe
dolphin	pistol	rifle
hawk	catapult	slingshot
beets	bazooka	cannon
raft	spear	sword
belt	bomb	missile
scooter	chicken	duck
jar	budgie	parakeet
dunebuggy	canary	finch
catapult	hawk	eagle
tomahawk	turkey	goose
van	caribou	moose
pistol	dolphin	whale
dresser	pineapple	coconut
toaster	prune	plum
cart	squash	pumpkin
chandelier	beans	peas
bazooka	beets	radish

APPENDIX B
Prime-target pairs from Experiment 2

Prime			
Less similar/more typical	More similar/less typical	Target	Dominant Category
sparrow	eagle	hawk	bird
robin	parakeet	budgie	bird
vulture	duck	chicken	bird
apple	plum	prune	fruit
peach	coconut	pineapple	fruit
corn	pumpkin	squash	vegetable
carrot	radish	beets	vegetable
cucumber	peas	beans	vegetable
toast	waffle	pancake	food
strudel	capcake	muffin	food
pepper	nutmeg	cinnamon	spice
ocean	stream	creek	body of water
shirt	bra	camisole	clothing
torch	lamp	chandelier	light source
silk	denim	corduroy	fabric
tuba	flute	clarinet	musical instrument
rake	hoe	shovel	gardening tool
gun	missile	bomb	weapon