4-24-2009

Conceptual Hierarchies in a Flat Attractor Network: Dynamics of Learning and Computations

Christopher M. O'Connor

George S. Cree

Ken McRae

Conceptual Hierarchies in a Flat Attractor Network: Dynamics of Learning and Computations

Christopher M. O'Connor,[a] George S. Cree,[b] Ken McRae[a]

[a]*Department of Psychology, University of Western Ontario, London*
[b]*Department of Psychology, University of Toronto Scarborough, Toronto*

Abstract

The structure of people's conceptual knowledge of concrete nouns has traditionally been viewed as hierarchical (Collins & Quillian, 1969). For example, superordinate concepts (*vegetable*) are assumed to reside at a higher level than basic-level concepts (*carrot*). A feature-based attractor network with a single layer of semantic features developed representations of both basic-level and superordinate concepts. No hierarchical structure was built into the network. In Experiment and Simulation 1, the graded structure of categories (typicality ratings) is accounted for by the flat attractor network. Experiment and Simulation 2 show that, as with basic-level concepts, such a network predicts feature verification latencies for superordinate concepts (*vegetable* <is nutritious>). In Experiment and Simulation 3, counterintuitive results regarding the temporal dynamics of similarity in semantic priming are explained by the model. By treating both types of concepts the same in terms of representation, learning, and computations, the model provides new insights into semantic memory.

Keywords: Superordinate concepts; Attractor networks; Temporal dynamics; Semantic memory

1. Introduction

When we read or hear a word, a complex set of computations makes its meaning available. Some words refer to a set of objects or entities in our environment corresponding to basic-level concepts such as *chair*, *hammer*, or *bean*, and thus refer to this level of information (Brown, 1958). Others refer to more general superordinate classes, such as *furniture*, *tool,* and *vegetable*, which encompass a wider range of possible referents. The goal of this article is to use a feature-based attractor network to provide insight into

how concepts at multiple ''levels'' might be learned, represented, and computed using an architecture that is not hierarchical.

A large body of research has implicated distinct treatment of basic-level and superordinate concepts. People are generally fastest to name objects at the basic level (Jolicoeur, Gluck, & Kosslyn, 1984), and participants in picture-naming tasks tend to use basic-level labels (Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976). Murphy and Smith (1982) have demonstrated similar effects with artificial categories. In ethnobiological studies of prescientific societies, the basic level (genus) is considered the most natural level of classification in folk taxonomic structures of biological entities (Berlin, Breedlove & Raven, 1973), and it emerges first in the evolution of language (Berlin, 1972). In addition, the time course of infants' development of representations for basic-level and superordinate concepts appears to differ (Rosch et al., 1976), with superordinates learned earlier than basic-level concepts (Mandler, Bauer, & McDonough, 1991; Quinn & Johnson, 2000). Complementary to this finding, during the progressive loss of knowledge in semantic dementia, basic-level concepts often are affected prior to superordinates (Hodges, Graham, & Patterson, 1995; Warrington, 1975). Thus, the way in which a concept is acquired, used, and lost depends partly on its specificity. Such differences have motivated semantic memory models in which basic-level and superordinate concepts are stored transparently at different levels of a hierarchy.

1.1. Hierarchical network models

Collins and Quillian's (1969) hierarchical network model was the first to capture differences between superordinate and basic-level concepts. They argued that concepts are organized in a taxonomic hierarchy, with superordinates at a higher level than basic-level concepts, and subordinate concepts at the lowest level. Features (*<is green>*) are stored at concept nodes, and the relations among concepts at different levels are encoded by ''is-a'' links. A central representational commitment was cognitive economy, so that features should be stored only at the highest node in the hierarchy for which they were applicable to all concepts below. An important processing claim was that it takes time to traverse nodes and to search for features within nodes. Collins and Quillian presented data supporting both cognitive economy and hierarchical representation. Given the model's successes, differences between superordinate and basic-level concepts were thought to be characterized, parsimoniously, and intuitively, by their location in a mental hierarchy.

Collins and Loftus (1975) extended this model in the form of spreading activation theory to account for some of its limitations. First, in a strict taxonomic hierarchy, a basic-level concept can have only one superordinate (Murphy & Lassaline, 1997). This proves problematic for many concepts; for example, *knife* can be a *weapon*, *tool,* or *utensil*. Collins and Loftus abandoned a strict hierarchical structure, allowing concept nodes from any level to be connected to any other. Second, Collins and Quillian's (1969) model was not designed to account for varying goodness, or typicality, of the exemplars within a category (e.g., people judge *carrot* to be a better example of a *vegetable* than is *pumpkin*). Numerous studies have used typicality ratings to tap people's knowledge of the graded structure of categories,

showing that it systematically varies across a category's exemplars (Rosch & Mervis, 1975). Collins and Loftus introduced a special kind of weight between basic-level and superordinate nodes (criteriality) to reflect typicality. This theory has been implemented computationally (Anderson, 1983), and it provides a comprehensive descriptive account of a large body of data (Murphy & Lassaline, 1997).

One limitation of these models, however, is that no mechanism has been described that determines which nodes are interconnected and what the strengths are on the connections. Without such a mechanism, the models may be unfalsifiable. This limitation has motivated researchers to instantiate new models in which weights between units are learned, and representations are acquired through exposure to structured inputs and outputs.

1.2. Connectionist models of semantic memory

Most computational investigations of natural semantic categories conducted in the past 25 years have been in the form of distributed connectionist models, and they have focused mainly on basic-level concepts (Hinton & Shallice, 1991; McRae, 2004; Plaut, 2002; Vigliocco, Vinson, Lewis, & Garrett, 2004). This focus is reasonable given the psychologically privileged status of the basic level. Consequently, when using models in which word meaning is instantiated across a single layer of units, as is typical of most connectionist models, it is not immediately obvious how to represent both basic-level and superordinate concepts.

A few connectionist models have addressed this issue. Hinton (1981, 1986) provided the first demonstration that they could code for superordinate-like representations across a single layer of hidden units from exposure to appropriately structured inputs and outputs. McClelland and Rumelhart (1985) showed that connectionist systems could develop internal representations, stored in a single set of weights, for both exemplar-like representations of individuals, and prototype-like representations of categories. Recently, Rogers and McClelland (2004, 2005) have extended this work to explore a broader spectrum of semantic phenomena.

The original aim of the Rogers and McClelland (2004) framework, as first instantiated by Rumelhart (1990) and Rumelhart and Todd (1993), was to simulate behavioral phenomena accounted for by the hierarchical network model. The model consists of two input layers, the item and relation layers, which correspond to the subject noun (*canary*) and relation (can/isa) in a sentence used for feature or category verification (''A canary can fly'' or ''A canary is a bird''). Each item layer unit represents a perceptual experience with an item in the environment (e.g., a particular *canary*). The relation layer units encode the four relations (has, can, is, ISA) used in Collins and Quillian (1969). The output (attribute) layer represents features of the input items. When the trained model is presented with *canary* and *has* as inputs, it outputs features such as <wings> and <feathers>, simulating feature verification. Rogers and McClelland also included superordinate (*bird*) and basic-level (*canary*) labels as output features. Thus, when presented with *canary* and *isa* as inputs, the model outputs *bird* and *canary*, simulating category verification.

Rogers and McClelland (2004) demonstrated that their model developed representations across hidden layer units that resembled superordinate representations. For example, if *canary* was presented as input and the model activated the superordinate node (*bird*), but not the basic-level node, this indicated that the model's representation for *canary* more closely resembled that of a superordinate than a basic-level concept. This pattern of results occurred only under circumstances where the model was unable to discriminate among individual items (e.g., *canary* and *robin*), as when the model was ''lesioned'' by adding noise to the weights or at points during training. Using this and other versions of the model, Rogers and McClelland (2004) provided insights into, most notably, patterns of impairment in dementia (Warrington, 1975) and numerous developmental phenomena (Gelman, 1990; Macario, 1991; Rosch & Mervis, 1975).

1.3. A new approach to distributed representation of superordinate concepts

We implement our theory in a distributed attractor network with feature-based semantic representations derived from participant-generated norms, and we provide both qualitative and quantitative tests. Our aim is to demonstrate that in treating concepts of different specificity identically in terms of the assumptions underlying learning and representation, one can capture the structure, computation, and temporal dynamics of basic-level and superordinate concepts. Our model extends those we have used to examine basic-level phenomena (Cree, McNorgan, & McRae, 2006; McRae, de Sa, & Seidenberg, 1997), and they borrow heavily from pioneering work in this area (Hinton, 1981, 1986; Masson, 1991; McClelland & Rumelhart, 1985; Plaut & Shallice, 1994; Rumelhart, 1990).

An important commonality between our model and that of Rogers and McClelland (2004) is that both depend on the statistical regularities among objects and entities in a human's environment (semantic structure) for shaping semantic representations. Observing the same features across repeated experiences with some entity or class of entities guides the abstraction of a coherent concept from perceptual experience (Randall, 1976; Rosch & Mervis, 1975; Smith, Shoben, & Rips, 1974). Category cohesion, the degree to which the semantic representations for a class of entities tends to overlap or hold together, shapes the specificity of a concept; the less cohesive the set of features that are consistently paired with a concept across instances, the more general the concept is, on average. Also relevant are feature correlations, the degree to which a pair of features co-occurs across multiple entities (e.g., something that <has a beak> also tends to <have wings>). Finally, regularities in labeling concepts at both the basic and superordinate levels play a key role in learning.

The present research extends Rogers and McClelland's (2004) approach in two important respects: by using a model that computes explicit feature-based superordinate representations, and by incorporating temporal dynamics in the model's computations. Rogers and McClelland were not primarily concerned with constructing a network that developed representations for superordinate concepts per se. In contrast, we simulated the learning of superordinate and basic-level terms in the following manner. On each learning trial, a concept's

label (name) was input, and it was paired with semantic features representing an instance of that concept in the environment. This simulates a central way in which we learn word meaning, through reading or hearing a word while the mental representation of its intended referent is active. For example, a parent might point to the neighbor's poodle and say ''dog.'' This labeling practice can be applied equivalently to basic-level and superordinate concepts. For example, people apply superordinate labels when referring to groups of entities (''I ate some *fruit* for breakfast''), physically present objects (''Pass me that *tool*''), or to avoid repetition in discourse (''She jumped into her car and backed the *vehicle* out of the driveway''). Thus, each superordinate learning trial consists of a superordinate label paired with an instance of that class. For example, the model might be presented with the word ''tool'' in conjunction with the semantic features of *hammer* on one trial, the features of *wrench* on another, and the features of *screwdriver* on yet another. In contrast, for basic-level concepts, the model is presented with consistent word-feature pairings. For example, the word ''hammer'' is always paired with the features of *hammer*.

In line with our training regimen, a number of studies of conceptual development support the idea that the connections established between a label and the corresponding set of perceptual instances are important in shaping the components of meaning activated when we read or hear a word (Booth & Waxman, 2003; Fulkerson & Haaf, 2003; Waxman & Markow, 1995). Plunkett, Hu, and Cohen (2007) presented 10-month-old infants with artificial stimuli that, based on their category structure, could be organized into a single category or into two categories. When the familiarization phase did not include labels, the infants abstracted two categories. When infants were provided with two labels consistent with this structure, the results were equivalent to the no-label condition. However, when infants were given two pseudo-randomly assigned labels, concept formation was disrupted. Crucially, presenting a single label for all stimuli resulted in the infants forming a single, one might say superordinate, representation despite the natural tendency for them to create two categories. These results demonstrate the importance of the interaction between labels and semantic structure and, in particular, how labeling leads to the formation of concepts at different ''levels.''

The second important way in which our modeling differs from Rogers and McClelland's (2004) is that our model incorporates processing dynamics. Rogers and McClelland used a feedforward architecture in which activation propagates in a single direction (from input to output). Because their model contained no feedback connections, directly investigating the time course of processing was not possible. Rogers and McClelland acknowledged the omission of recurrent connections to be a simplification and assumed that such connections are present in the human semantic system. In contrast, a primary goal of our research is to demonstrate the ability of a flat connectionist network to account for behavior that unfolds over time, such as feature verification and semantic priming. Therefore, we used an attractor network, a class of connectionist models in which recurrent connections enable settling into a stable state over time, allowing us to investigate the time course of the computation of superordinate and basic-level concepts.

1.4. Overview

We describe the model in Section 2. In Section 3, we demonstrate that it develops representations for superordinate and basic-level concepts that fit with intuition and that capture superordinate category membership. Section 4 presents quantitative demonstrations of the relations among these concept types by simulating typicality ratings. In Section 5, the model's representations of superordinates are investigated using a speeded feature verification task. In Section 6, we use the contrast between the model's superordinate and basic-level representations, in conjunction with its temporal dynamics, to provide insight into the counterintuitive finding that superordinates equivalently prime high and low typicality basic-level exemplars. This result is inconsistent with all previous theories of semantic memory because those frameworks predict that the magnitude of such priming effects should reflect prime-target similarity. Interestingly and surprisingly, the model accounts for all of these results using a flat representation, that is, without transparently instantiating a conceptual hierarchy.

2. The model

We first present the derivation of the features used to train the basic-level and superordinate concepts, followed by the model's architecture. The manner in which the model computes a semantic representation from word form, and the training regime, is then described.

2.1. Concepts

2.1.1. Basic-level concepts

The semantic representations for the basic-level concepts were taken from McRae, Cree, Seidenberg, and McNorgan's (2005) feature production norms (henceforth, ''our norms''). Participants were presented with basic-level names, such as *dog* or *chair*, and were asked to list features. Each concept was presented to 30 participants, and any feature that was listed by five or more participants was retained. The norms consist of 541 concepts that span a broad range of living and nonliving things. This resulted in a total of 2,526 features of varying types (Wu & Barsalou, in press), including external and internal surface features (*bus* <is yellow>, *peach* <has a pit>), function (*hammer* <used for pounding nails>), internal and external components (*car* <has an engine>, *octopus* <has tentacles>), location (*salmon* <lives in water>), what a thing is made of (*fork* <made of metal>), entity behaviors (*dog* <barks>), systemic features (*ox* <is strong>), and taxonomic information (*violin* <a musical instrument>).

All taxonomic features were excluded for two reasons, resulting in 2,349 features. First, features that describe category membership are arguably different from those that describe parts, functions, and so on. Second, it could be argued that including taxonomic features in the model would be equivalent to providing hierarchical information, which we wished to avoid.

2.1.2. Superordinate concepts

The goal was to have the model learn superordinates via its experience with basic-level exemplars. Therefore, the superordinate features such as carrot <a vegetable> were used to establish categories and their exemplars. These features indicated the category (or categories) to which the norming participants believed each basic-level concept belonged (if any). Using a procedure similar to that of Cree and McRae (2003), a basic-level concept was considered a member of a superordinate category if at least two participants listed the superordinate feature for that concept. A superordinate was used if it was listed for more than 10 basic-level concepts. The goal of using this criterion was to include a reasonably large, representative sample of exemplars for each superordinate. The sole exception was plant, which was excluded because only five of the 18 exemplars were not fruits or vegetables, and therefore the sample was not representative. These criteria resulted in 20 superordinates, ranging from 133 exemplars (animal) to 11 (fish). The number of superordinates with which a basic-level concept was paired ranged from 0 (e.g., ashtray, key) to 4 (e.g., cat is an animal, mammal, pet, and predator). The resulting 611 superordinate–exemplar pairs are presented in Appendix A.

2.2. Architecture

The network consisted of two layers of units, wordform and semantics, described below (see Fig. 1). All 30 wordform units were connected unidirectionally to the 2,349 semantic feature units. The wordform units were not interconnected. All semantic feature units were fully interconnected (with no self-connections). Although all connections were unidirectional, semantic units were connected in both directions (i.e., two connections) so that activation could pass bidirectionally between each pair of feature units.

2.2.1. Wordform input layer

Each basic-level and superordinate concept was assigned a three-unit code such that turning on (activation = 1) the three units denoting a concept name and turning off (activation = 0) the remaining 27 units can be interpreted as presenting the network with the spelling or sound of the word. Of the 4,060 possible input patterns, 541 unique three-unit combinations were assigned randomly to the basic-level concepts, and 20 to the superordinates. Random overlapping wordform patterns were assigned because there is generally no
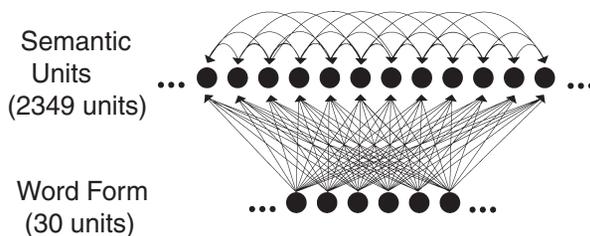

Fig. 1. Model architecture.

systematic mapping from wordform to semantics in English monomorphemic words, and many concept names overlap phonologically and orthographically.

2.2.2. Semantic output layer

Each output unit corresponded to a semantic feature from our norms. Thus, concepts were represented as patterns of activation distributed across semantic units. Because semantic units were interconnected, the model naturally learned correlations between feature pairs. Thus, if two features co-occur in a number of concepts (as in <has wings> and <has feathers>), then if one of them is activated, the other will also tend to be activated by positive weights between them. Alternatively, if two features tend not to co-occur (<has feathers> and <made of metal>), then if one is activated, the other will tend to be deactivated.

The use of features as semantic representations is not intended as a theoretical argument that the mental representation of object concepts exists literally in terms of lists of verbalizable features. However, when participants perform feature-listing tasks, they make use of the holistic representation for concepts that they have developed through multisensory experience with things in the world (Barsalou, 2003). Thus, this empirically based approximation of semantic representation provides a window into people's mental representations that captures the statistical regularities among object concepts. In addition, these featural representations provide a parsimonious and interpretable medium for computational modeling and human experimentation.

2.3. Computation of word meaning

To compute word meaning, a three-unit concept name was activated and remained active for the duration of the computation. Semantic units were initialized to random values between .15 and .25. Activation spread from each wordform to each semantic unit, as well as between each pair of semantic units. Input to a semantic unit was computed as the activation of a sending unit multiplied by the weight of the connection from that sending unit. The net input x_j (at tick t) to $unit_j$ was then computed according to Eq. 1:

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

where s_i is the activation of $unit_i$ sending activation to $unit_j$, and w_{ji} is the weight of the connection between $unit_i$ and $unit_j$. τ (tau) is a constant between 0 and 1 used to denote the duration of each time tick (.2 in our simulations). As such, $x_j^{[t-\tau]}$ denotes net input to $unit_j$ at the previous time tick. Each of these time ticks is a subdivision of a time step, and it consists of passing activation forward one step. Time ticks are used to discretize and simulate continuous processing between time steps (Plaut, McClelland, Seidenberg, & Patterson, 1996). The net input to $unit_j$ at time t is converted to an activation value ($a_j^{[t]}$) according to the sigmoidal activation function presented in Eq. 2, where $x_j^{[t]}$ is the net input to $unit_j$ from Eq. 1 (at time t).

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

Activation propagated for four time steps, each of which was divided into five ticks, for a total of 20 time ticks. After the network had been fully trained, it settled on a semantic representation for a word in the form of a set of activated features at the semantic output layer.

2.4. Training

Prior to training, weights were set to random values between −.05 and .05. For each training trial, the model was presented with a concept's wordform and activation accrued over four time steps (ticks 1–20). For the last two time steps (ticks 11–20), the target semantic representation was provided, and error was computed. This allows a concept's activation to accumulate gradually because the training regime does not constrain the network to compute the correct representation until tick 11. The cross-entropy error (CEE) metric was used because it is more suitable than are the more frequently used squared-error metrics for two reasons. First, during training, features can be considered as being either present (on) or absent (off), and intermediate states are understood as the probability that each feature is present. Thus, the output on the semantic layer represents a probability distribution that is used in computing cross entropy (Plunkett & Elman, 1997). Second, this error metric is advantageous for a sparse network such as this one—where only five to 21 of the 2,329 feature units should be on for the basic-level concepts—because it produces large error values when a unit's activation is on the wrong side of .5. One potential way for our sparse network to reduce error is to turn off all units at the semantic layer. Thus, punishing the network for incorrectly turning units off allows the model to more easily change the states of those units. CEE (*E*), averaged over the last two time steps (10 ticks) was computed as in Eq. 3:

$$E = \tau \frac{\sum_{t=10}^{19} \sum_{j=0}^{2348} d_j \ln(a_j) + (1 - d_j) \ln(1 - a_j)}{10} \tag{3}$$

where d_j is the desired target activation for unit$_j$, and a_j is the unit's observed activation.

After each training epoch, in which every concept was presented, weight changes were calculated using the continuous recurrent backpropagation-through-time algorithm (Pearlmutter, 1995). The learning rate was .01 and momentum (.9) was added after the first 10 training epochs (Plaut et al., 1996). Training and simulations were performed using Mikenet version 8.02.

2.4.1. Basic-level concepts
The basic-level concepts were trained using a one-to-one mapping. That is, for each basic-level training trial, the network learned to map its wordform to the same set of semantic features. Error correction was scaled by familiarity. This was based on familiarity ratings in which participants were asked how familiar they were with ''the thing to which the word refers'' on a 9-point scale, where 9 corresponded to ''extremely familiar'' and 1 to ''not

familiar at all'' (McRae et al., 2005). The scaling parameter was applied to the error computed at each unit in the semantic layer, and it was calculated using Eq. 4:

$$S_b = \frac{.9 \times \text{fam}_b}{\max(\text{fam})} \tag{4}$$

where S_b is the scaling parameter applied to a basic-level concept, fam_b is the familiarity of that concept, and $\max(\text{fam})$ is the maximum familiarity rating across all basic-level concepts. Thus, familiar concepts exerted a greater influence on the weight changes than did unfamiliar concepts, simulating people's differential experience with various objects and entities. The reason for including .9 is described in Section 2.4.3.

2.4.2. Superordinate concepts

Superordinate training was somewhat more complex. The operational assumption was that people learn the meaning associated with a superordinate label by experiencing that label paired with specific exemplars (note that if a superordinate label is used to refer to groups of objects, this manner of training also applies). Thus, superordinate concepts were trained using a one-to-many mapping. Rather than consistently mapping from a given wordform to a single set of features (as was done with basic-level concepts), the model was trained to map from a superordinate wordform to various exemplar featural representations on different trials. For example, when presented with the wordform for *vegetable*, on some learning trials it was paired with the features of *asparagus*, and on others those of *broccoli*, *carrot,* and so on. Over an epoch of training, the model was presented with the featural representations of all 31 *vegetable* exemplars, each paired with the wordform for *vegetable*. This process was performed for all 20 categories and all exemplars in each category.

We assume that, in reality, the process by which both types of concepts are learned is identical. Although basic-level concepts were trained in a one-to-one fashion, people experience individual apples, bicycles, and dogs, and these instances often are paired with basic-level labels (generated either externally or internally). If the model had learned basic-level concepts through exposure to individual instances, such as individual apples, we assume that it would develop representations similar to the ones with which it was presented in the training herein. This would occur because instances of basic-level concepts tend to overlap substantially in terms of features. For example, almost every *onion* <is round>, <tastes strong>, <has skin>, and so on. Therefore, if the model was presented with individual onions, these features would emerge from training with activations close to 1, due to the high degree of featural overlap (cohesion) among instances. The one-to-one mapping assumption was made, therefore, to simplify training.

The assumption also applies to training superordinates. The way that basic-level and superordinate concepts are learned is qualitatively identical. That is, in reality, superordinate concepts are also learned through exposure to instances. As such, if the model was presented with individual instances of onions, carrots, and beets along with the wordform for *vegetable*, it would presumably develop representations similar to those it learned using the present training regime, in which it was trained on basic-level representations.

2.4.3. Scaling issues

Three scaling issues concerning superordinate learning were considered. First, it was necessary to decide how often the network should associate each exemplar with its superordinate label. There are multiple possibilities. For example, exemplars that are more familiar could be paired with the superordinate label more frequently under the assumption that familiar exemplars exert a stronger influence in learning a superordinate than do unfamiliar exemplars. Conversely, it could be argued that exemplars that are less familiar are more likely to be referenced with their superordinate label because people have greater difficulty retrieving the concepts' basic-level names and use the superordinate name instead. Although both of these possibilities (and others) are reasonable, there is no research to suggest that any option is more valid than another. Therefore, we used the training regime that relied on the fewest assumptions; we presented each exemplar equiprobably. For example, if a category had 20 exemplars, each was associated with the superordinate label 5% of the time.

A crucial point here is that pairing a superordinate label equally frequently with the semantic representation of each of its exemplars means that typicality was not trained into the model. This is critical because behavioral demonstrations of graded structure are a major hallmark of category-exemplar relations, and capturing the influence of typicality is an important component of the experiments and simulations presented herein.

The second issue concerns how often to train superordinates in relation to each other. Because of the general nature of superordinate concepts, it seemed inappropriate to use subjective familiarity as an estimate. That is, it is unclear what it means to ask someone how familiar they are with the things to which *furniture* refers. The most reasonable solution was to use word frequency under the assumption that the amount that people learn about superordinate concepts varies with the frequency with which its label is applied. Therefore, superordinate concepts were frequency weighted in a fashion similar to the weighting of the basic-level concepts by familiarity. Superordinate name frequency was measured as the natural logarithm of the concept's frequency taken from the British National Corpus, ln(BNC). The scaling parameter applied to the error measure was ln(BNC) of a superordinate name (summed over singular and plural usage), divided by the sum over all 20 superordinates.

The final issue concerns how to scale superordinate relative to basic-level training. Wisniewski and Murphy (1989) counted instances of basic-level and superordinate names in the Brown Corpus (Francis & Kucera, 1982). They found that 10.6% (298/2,807) were superordinate names and 89.4% (2509/2,807) were basic level. In mother–child discourse, Lucariello and Nelson (1986) found that mothers used superordinate labels 11.4% of the time (142/1,244) and basic-level labels 88.6% of the time (1,102/1,244). Therefore, we trained superordinates 10% of the time, and basic level concepts 90%. This was simulated by scaling the error at each semantic unit, which is why .9 appears in Eq. 4, and .1 in Eq. 5. Thus, basic-level concepts had a greater influence than superordinate concepts on the changes in the weighted connections during training.[1]

To summarize, the scaling parameter (S_s) applied to the error computed at each semantic unit was calculated for each superordinate label-exemplar features pair using Eq. 5:

$$S_s = \frac{.1 \times \ln(\text{BNC}_s)}{n_s \times \sum_{k=1}^{20} \ln(\text{BNC}_k)} \tag{5}$$

where BNC_s is the frequency of the superordinate name from the British National Corpus, and n_s is the number of exemplars in the superordinate's category. The scaling parameter included n_s so that all pairings of a single superordinate label with its exemplars' semantic representations summed to the equivalent of a single presentation of the superordinate concept.

2.4.4. Completion of training

Because the model was allowed to develop its own superordinate representations, they could not be used to determine when to terminate training. Ordinarily, training is stopped when a model has learned the patterns to some prespecified criterion. In this case, the ''correct'' featural representation for the superordinates was unknown because they were trained in a one-to-many fashion. However, the target features for the basic-level concepts were known. Therefore, training stopped when the model had successfully learned them; that is, when 95% of the features that were intended to be on (activation = 1) had an activation level of .8 or greater (across all 541 concepts). This was achieved after 150 training epochs.

By training the network until a single point in time where all superordinate and basic-level concepts had attained a static and stable semantic representation, we are not arguing that humans possess a static representation for either type of concept. On the contrary, we believe that concepts constantly develop and vary dynamically over time such that they are sensitive to context and aggregated experience (Barsalou, 2003). For these reasons, as well as the fact that knowledge is stored in the weights in networks such as these, we use the term ''computing meaning,'' rather than, for example, ''accessing meaning,'' throughout this article.

3. Superordinate representations and category membership

3.1. Superordinate representations

To verify that the model developed reasonable representations for the superordinate concepts, the network was presented with the three-unit wordform for each superordinate and was allowed to settle on a semantic representation over 20 time ticks. As one example, the features of *vegetable* with activation levels greater than .2 after 20 time ticks are presented in Table 1. Interestingly, although one feature is activated close to 1 (<is edible>), most are only partially activated. For example, <grows in gardens> and <eaten in salads> have activation levels of .56 and .39, respectively. This pattern of midlevel activations is characteristic of all of the computed superordinate representations. Contrast this with the activation pattern for *celery*, also presented in Table 1. The semantic units of all basic-level concepts had activation levels close to 1 or 0. For example, of all *celery* features with activations greater than .2, the lowest were <has stalks> and <tastes good> at .89. This difference in

Table 1
Computed network representations for *vegetable* and *celery*

Concept	Feature	Activation
Vegetable	<is edible>	.79
	<grows in gardens>	.56
	<is green>	.56
	<eaten by cooking>	.48
	<is round>	.48
	<eaten in salads>	.39
	<is nutritious>	.38
	<is small>	.34
	<is white>	.33
	<tastes good>	.33
	<has leaves>	.31
	<is crunchy>	.28
	<grows in the ground>	.24
Celery	<is green>	.98
	<grows in gardens>	.97
	<is crunchy>	.96
	<is nutritious>	.94
	<has leaves>	.93
	<eaten in salads>	.92
	<is edible>	.91
	<is long>	.91
	<is stringy>	.91
	<tastes bland>	.91
	<eaten with dips>	.90
	<has fiber>	.90
	<has stalks>	.89
	<tastes good>	.89

activation patterns results from three factors underlying the semantic structure of conceptual representations: feature frequency, category cohesion, and feature correlations, all of which influence learning.

3.1.1. Feature frequency

As would be expected, the activation levels of superordinate features depend on the number of exemplars possessing each feature. When the model (and presumably a human as well) is presented with the label for a superordinate, such as *vegetable*, it is not always paired with the same set of features. Therefore, features that appear in many (or all) members of the category, such as <is edible>, are highly activated. Those that appear in some members like <grows in gardens> have medium levels of activation, and those appearing in a few concepts, like <is orange>, have low activation. Thus, not all superordinate concepts have only a few features with high activations. For example, *bird* has a number of features with activations greater than .85, such as <has a beak>, <has feathers>, <has wings>, and <flies>. This is consistent with the suggestion that *bird* may actually be a basic-level concept.

3.1.2. Category cohesion

Many other concepts, like *fish* or *tree*, can also be argued to function as basic-level as well as superordinate concepts (Rosch et al., 1976). The exemplars of these categories share many features (e.g., basically all birds include <has feathers>, <has a beak>, and <has wings>) and differ in few features, such as their color. The patterns of features for these concepts are consequently highly cohesive, producing strongly activated features for *bird*. In contrast, other superordinates such as *furniture* have few features that are shared by many exemplars (e.g., <made of wood>). Thus, the exemplars of these categories are not cohesive in terms of overlapping features because they differ in many respects (e.g., color, external components, function). Thus, category cohesion has direct implications for the conceptual representations that the models learn to compute. For example, *furniture* included few strongly activated features.

3.1.3. Feature correlations

For the present purposes, features were considered correlated if they tend to co-occur in the same basic-level concepts (McRae et al., 1997). For example, various birds include both <has wings> and <has feathers>. Attractor networks naturally learn these correlations, and they play an important role in computing word meaning. People also learn these distributional statistics implicitly by interacting with the environment, and these statistics influence conceptual computations (McRae, Cree, Westmacott, & de Sa, 1999). In the present model, features that are mutually correlated activate one another during the computation of superordinate concepts (and basic-level concepts as well). Therefore, feature correlations, particularly across exemplars of a superordinate category, strongly influence what features are activated for a superordinate.

3.2. Delineating members from nonmembers

We begin with a simulation that addresses whether the network can delineate between category members versus nonmembers. We computed the representations for all superordinate and basic-level concepts by inputting the wordform for each and then letting the network settle. We calculated the cosine between the network's semantic representations for every superordinate-basic level pair, thus providing a measure of similarity, as in Eq. 6.

$$\cos(\vec{x}, \vec{y}) = \frac{\sum_{i=1}^{2349} x_i y_i}{\sqrt{\sum_{i=1}^{2349} x_i^2} \sqrt{\sum_{i=1}^{2349} y_i^2}} \tag{6}$$

In Eq. 6, $\vec{x}$ is the vector of feature activations for a superordinate, $\vec{y}$ is the vector for an exemplar, and x_i and y_i are the feature unit activations.

We sorted the basic-level concepts in terms of descending similarity to each superordinate. For testing the model, an exemplar was considered a category member if at least 2 of 30 participants in our norms provided the superordinate category as a feature for the basic-level concept (the same criterion used to generate the superordinate-basic level pairs).

Our measure of the extent to which the model can delineate categories was the number of basic-level concepts that were ranked in the top *n* in terms of similarity to each superordinate, where *n* is the number of category members (according to the 2 of 30 norming participants criterion). For example, there are 39 *clothing* exemplars, so we counted the number of them that fell within the 39 exemplars that were most similar to *clothing*. This provides a reasonably conservative measure of the network's performance because the criterion is a liberal estimate of category membership (i.e., it includes those exemplars residing at the fuzzy boundaries, McCloskey & Glucksberg, 1978).

In general, the network captures category membership extremely well (see Table 2). *Clothing* and *bird* were perfect in that all category members had higher cosines than did any nonmembers. There was only a single omission for five superordinates, and the network's errors were quite reasonable: *furniture* (*lamp*), *fish* (*shrimp* preceded *guppy*), *container* (*dish* preceded *freezer*), *vehicle* (*canoe* was 29th, whereas *wheelbarrow* intruded at 23rd; there were 27 vehicles), and *musical instrument* (*bagpipe* was one slot lower than it should have been). There were two omissions for *appliance* (*radio* and *telephone* came after *corkscrew* and *colander*), and *insect* (*caterpillar* and *flea* came after two birds, *oriole* and *starling*). Four superordinates contained three omissions: *animal* (missed *crab*, *python*, and surprisingly, *bull*, while intruding *grasshopper*, *housefly*, and *sardine*, which are all technically animals); *carnivore* (*dog*, *alligator*, and *porcupine* lay outside the cutoff); *herbivore* (missed *turtle*, *giraffe*, and *grasshopper*, although two of three false alarms are actually herbivores,

Table 2
The network's prediction of category membership

Category	Number of Exemplars	Number Within Criterion	Percent Correct
Furniture	15	14	93
Appliance	14	12	86
Weapon	39	33	85
Utensil	22	19	86
Container	14	13	93
Clothing	39	39	100
Musical instrument	18	17	94
Tool	34	25	74
Vehicle	27	26	96
Fruit	29	27	93
Bird	39	39	100
Insect	13	11	85
Vegetable	31	27	87
Fish	11	10	91
Animal	133	130	98
Pet	22	17	77
Mammal	57	51	89
Carnivore	19	16	84
Herbivore	18	15	83
Predator	17	14	82

donkey, and *chipmunk*, but not *mink*); and *predator* (*tiger*, *hyena*, and *alligator* fell below *crocodile*, *buzzard*, and *hare*, although why participants listed predator for *alligator* but not for *crocodile* is a bit of a mystery).

Other categories included mammal, for which there were five errors, only one of which is clearly a mammal (*mole*), and *pet*, for which category membership was captured for only 17 of 22 exemplars. There was a bit of *fruit/vegetable* confusion, although this is common with people as well. For *fruit*, *pickle* and *peas* preceded *pumpkin* and *rhubarb*, and for *vegetable*, *garlic*, *corn*, *pumpkin*, and *pepper* were preceded by some exemplars that are clearly fruits, *pear*, *strawberry*, and *blueberry*. Finally, there were also some *tool/weapon* confusions. There were six items that were erroneously included as weapons, although half of them could be used as such (*scissors*, *spade*, and *rake*; see one of numerous movies for examples). The omitted exemplars were *catapult*, *whip*, *stick*, *stone*, *rock*, and *belt*, all of which are atypical weapons. There were likewise *weapon/tool* confusions at the boundary of *tool*.

In summary, the network's representations clearly capture category membership. The errors primarily reflect our liberal criterion for category membership, occurring at the fuzzy boundaries, and are reasonable in the vast majority of cases. In Experiment 1 and Simulation 1, we investigate whether the network can account for a related, but somewhat more fine-grained measure, graded structure within those categories.

4. Experiment 1 and Simulation 1: Typicality ratings

Categories exhibit graded structure in that some exemplars are considered to be better members than others (Rips, Shoben, & Smith, 1973; Rosch & Mervis, 1975; Smith et al., 1974). Typicality ratings have been used extensively as an empirical measure of this structure. Thus, it is important that our model accounts for them. Participants provided typicality ratings for the superordinate and basic-level concepts on which the network was trained. This task was simulated, and the model's ability to predict the behavioral data was tested. Family resemblance was used as a baseline for assessing the model's performance. Family resemblance is known to be an excellent predictor of behavioral typicality ratings (Rosch & Mervis, 1975). Lastly, interesting insights are gained by considering the successes and shortfalls of the model.

4.1. Experiment 1

4.1.1. Method
4.1.1.1. Participants: Forty-two undergraduate students at the University of Western Ontario participated for course credit, 21 per list. In all studies reported herein, participants were native English speakers and had normal or corrected-to-normal visual acuity.

4.1.1.2. Materials: Experiment 1 was conducted prior to the current project and included other superordinate categories. Typicality ratings were collected for all categories used by

Cree and McRae (2003). Thus, there were 33 superordinate categories and 729 nonunique exemplars (many exemplars appeared in multiple categories). Because 729 rating trials were deemed to be too many for a participant to complete, there were two lists. The first list consisted of 17 categories and 373 basic-level concepts, and the second consisted of 16 categories and 356 basic-level concepts. Both lists contained concepts from the living and nonliving domains.

4.1.1.3. Procedure: Instructions were presented on a Macintosh computer using Psy-Scope (Cohen, MacWhinney, Flatt, & Provost, 1993) and were read aloud to the participant. Participants were told they would see a category name as well as an instance of that category, and they were asked to rate how ''good'' an example each instance is of the category, using a 9-point scale which was presented on the screen for each trial. They were instructed that ''9 means you feel the member is a very good example of your idea of what the category is,'' and ''1 means you feel the member fits very poorly with your idea or image of the category (or is not a member at all).'' Participants were provided with an example and performed 20 practice trials using the category *sport*, with verbal feedback if requested. Categories were presented in blocks, with randomly ordered presentation of exemplars within a category. The typicality rating for each superordinate–exemplar pair was the mean across 21 participants. No time limit was imposed, and participants were instructed to work at a comfortable pace.

4.2. Simulation 1

4.2.1. Method
4.2.1.1. Materials: The items were the 611 superordinate–exemplar pairs on which the model was trained.

4.2.1.2. Procedure: To simulate typicality rating, we first initialized the semantic units to random values between .15 and .25 (as in training). The basic-level concept's wordform (e.g., *celery*) was then activated and the network was allowed to settle for 20 ticks. The computed representation was then recorded, as if it was being held in working memory. The semantic units were reinitialized, the superordinate's wordform (e.g., *vegetable*) was presented, and the network was allowed to settle for 20 ticks. To simulate typicality ratings, we calculated the cosine between the network's computed semantic representations for each superordinate-basic-level pair as in Eq. 6. To test the model's ability to account for typicality ratings, the cosine between each superordinate–exemplar pair was correlated with participants' mean typicality ratings.

Family resemblance was calculated using the feature production norms and computed as in Rosch and Mervis (1975). For all exemplars of a superordinate, each feature from the norms received a score corresponding to the number of concepts in the category that possess that feature. The family resemblance of an exemplar within a category was the sum of the scores for all of the exemplar's features.

4.3. Results and discussion

Because positive correlations were predicted, and there was no reason to expect negative correlations, all reported *p*-values are based on a one-tailed distribution. As presented in Table 3, the Pearson product-moment correlation between the model's cosine similarity and the typicality ratings was significant for 13 of 20 categories, showing that the model was successful in simulating graded structure.[2] By comparison, family resemblance predicted typicality ratings for those same 13 categories. Thus, the predictive abilities of the model and family resemblance were comparable. Furthermore, all correlations between model cosine and family resemblance were significant except one (container, $p = .052$), and they ranged from .45 to .93. These correlations reflect the fact that the number of exemplars that possess a feature is a major contributor to the superordinates' representations in the model. That the correlations are not consistently extremely high reflects the fact that feature correlations influence learning and computations in the network, but not in the family resemblance measure. There were, however, a few categories that proved difficult for the model and family resemblance in terms of simulating graded structure.

Table 3
Correlations among superordinate/basic-level cosine from the model (Simulation 1),
family resemblance scores, and typicality ratings (Experiment 1)

Category	*N*	Cosine and Typicality	Family Resemblance and Typicality	Cosine and Family Resemblance
Furniture	15	.76**	.62**	.74**
Appliance	14	.69**	.73**	.92**
Weapon	39	.63**	.70**	.83**
Utensil	22	.50**	.52**	.73**
Container	14	.49*	.50*	.45
Clothing	39	.46**	.50**	.81**
Musical instrument	18	.44*	.54*	.93**
Tool	34	.38*	.38*	.69**
Vehicle	27	−.02	.18	.77**
Fruit	29	.73**	.69**	.90**
Bird	39	.62**	.49**	.70**
Insect	13	.55*	.69**	.77**
Vegetable	31	.47**	.51**	.89**
Fish	11	.38	.36	.91**
Animal	133	.12	.12	.52**
Pet	22	.08	.02	.92**
Mammal	57	.02	.20	.73**
Carnivore	19	.61**	.45*	.77**
Herbivore	18	−.14	.06	.57**
Predator	17	−.05	.21	.77**

Note: *$p < .05$, **$p < .01$.

4.3.1. Low variability

Fish was problematic because of little variability in the human typicality ratings of the *fish* exemplars (*SD* = .86) compared to other categories, such as *tool* (*SD* = 1.59) or *vegetable* (*SD* = 1.67). In addition, the typicality and cosine ratings of the *fish* were high (*M* = 7.58 and .71, respectively) compared, for example, to *tool* (*M* = 6.18 and .43) or *vegetable* (*M* = 6.57 and .49). The reason for this appears to be that people seem to know little that distinguishes individual types of fish. Therefore, given that people know only general features of *cod*, *mackerel*, *perch*, and so on, such as <has gills>, <has scales>, and <swims>, they rate all of these exemplars as typical because these features are shared by all fish. Given the limited variability, combined with the fact that this category includes only 11 exemplars in our norms, it is not surprising that it proved difficult for the model and for family resemblance in terms of predicting graded structure.

4.3.2. Liberal sampling

The *vehicle* category was also an issue for family resemblance and the model. This was surprising given that there were 27 exemplars, a reasonable number for the model to use in developing superordinate representations. However, many of the exemplars in this category are what might be considered atypical. Of the 27 vehicles, 17 were listed as a vehicle by fewer than five of the 30 participants in the feature production norms. The remaining 10 exemplars—*ambulance, bus, car, dunebuggy, jeep, motorcycle, scooter, tricycle, truck,* and *van*—are composed of road-worthy vehicles, suggesting that perhaps the true *vehicle* representation might be more similar to them.

To test this hypothesis, the semantic representation for *car* was treated as the superordinate representation for *vehicle*, and the typicality rating task was resimulated. The correlation between the new cosine measures in the model and the original typicality ratings for *vehicle* using 26 exemplars (i.e., excluding *car* itself) was .38, *p* < .05. This improvement in correlation (from −.02) indicates that the original *vehicle* representation is not entirely representative. It seems reasonable to assume that during learning in the real world, people simply do not refer to sailboats, canoes, and skateboards using *vehicle*, and therefore including these items equiprobably in the training phase was not particularly realistic.

4.3.3. Mammal and animal

These have been regarded as special cases. People often use *animal* as the superordinate of individual mammals more frequently than they use *mammal* (Rips et al., 1973). People also correctly verify statements such as ''A cow is an animal'' faster than ''A cow is a mammal,'' despite the fact that *mammal* would seem to be the more directly relevant superordinate. This unfamiliarity with *mammal* resulted in participants seemingly relying primarily on size for rating typicality, whereby large mammals were rated as more typical (e.g., *whale*, *M* = 7.5) and small mammals as less typical (e.g., *mouse*, *M* = 6.05). In contrast, the model's performance depended on overlapping features (such as <has 4 legs>, and <has fur>) in simulating typicality ratings (*whale* has a cosine of .40, whereas *mouse* has a cosine of .52).

Animal is unusual because it has been argued to have a number of senses (Deese, 1965; Rips et al., 1973). For example, *animal* can be considered in its scientific sense as being the

superordinate of *bird*, *mammal*, *insect*, and so on, whereas average people use *animal* to refer to the biological category of mammals. In support of this idea, when the model's representation for *mammal* was used to predict typicality ratings for the 133 animal exemplars, the correlation increased from .12 to .56, $p < .001$. The poor performance of the model and family resemblance is due to the fact that the representation for *animal* is undoubtedly influenced by the pairing of the *animal* label with many *birds*, *fish*, and *insects*.

4.3.4. Role-governed categories

The three other categories that were problematic for the model were *herbivore*, *pet*, and *predator*. The reason for this may be that these categories are not learned based on featural similarity. They may be better thought of as role-governed categories, categories defined by their role in a relational structure (Markman & Stilwell, 2001). For example, in the relation hunt(x, y), x plays the role of the hunter and y plays the role of the hunted. In this case, the category *predator* is defined by its role as the first argument, x, in the relation. This therefore defines the exemplars of *predator* based on their role in the relation (e.g., *alligator*, *cat*, and *falcon* hunt y). This category type also applies to *pet*, as the second argument in the relation domesticate(x, y). More interesting, however, are *carnivore* and *herbivore*. In these cases categorization is contingent on the second argument in the relation eat(x, y). In the above, the status of one argument was irrelevant (e.g., it did not matter *what* was hunted). In this case, *carnivore* applies to any x (e.g., *crocodile*) in the relation eat(x, y) where y is *flesh*, and *herbivore* applies to any x (e.g., *deer*) where y is *plants* (although we recognize these are slight oversimplifications).

Given the apparent inability of the model (or family resemblance) to predict typicality ratings for these categories, it appears that featural similarity is not generally useful for their classification. Consider, for example, that *birds*, *fish*, *insects*, and *mammals* can all be *herbivores*, *pets*, or *predators*. However, that the model was able to predict typicality ratings for *carnivore* emphasizes Markman and Stilwell's (2001) argument that role-governed and feature-based theories need not be independent. The primary reason that the model was successful with *carnivore* is that being a *carnivore* entails possessing a number of features that facilitate hunting and eating meat, like <has claws>, <has teeth>, and <is large>. Therefore, these features are correlated for *carnivore*, allowing a coherent representation to be formed.

In summary, the model predicts human typicality ratings, producing results that are roughly equivalent to family resemblance. In Simulation and Experiment 2, we tested the model in a somewhat more specific manner by investigating whether the degree to which specific features are activated by superordinate names can predict human feature verification latencies.

5. Experiment 2 and Simulation 2: Superordinate feature verification

In the model, features of superordinates are activated gradually as a superordinate is computed. Thus, this is the first model to introduce temporal dynamics into the computation of superordinate representations. In addition, the model activates a superordinate's features to

varying and generally intermediate degrees. These characteristics enabled the generation of testable predictions for a new task, superordinate feature verification.

Feature verification experiments with basic-level concepts have been used to gain insight into a number of aspects of concept-feature relations (Pecher, Zeelenberg, & Barsalou, 2003; Solomon & Barsalou, 2004). In some cases, the results have been simulated using attractor networks (Cree et al., 2006; McRae et al., 1999). In those simulations, we assumed that a feature's activation while a concept is being computed is monotonically related to human feature verification latency. Therefore, simulations were conducted by activating a concept's wordform and recording the activation of the relevant feature over time. Experiment and Simulation 2 investigated whether feature activation during the model's computation of superordinate concepts predicts superordinate-feature verification latencies using the identical method as was used with previous studies of basic-level concepts.

5.1. Experiment 2

5.1.1. Method
5.1.1.1. Participants: Twenty-six University of Western Ontario undergraduates participated in course credit. One was excluded because his/her mean decision latency was an extreme outlier, and two were excluded because their error rates were extreme outliers.

5.1.1.2. Materials: Eighteen of the 20 superordinates were used, 10 living and eight nonliving things. *Animal* was omitted because of insufficient variability in the activation levels of the features; no feature was activated greater than .5. *Musical instrument* was omitted because its name consists of two words, and therefore the time required to read it might overlap with the presentation of the feature, which would artificially lengthen decision latencies.

Fifty-four target trials were constructed by pairing each superordinate with three features (see Appendix B). Because regression analyses were the focus, to provide a suitable distribution, we used features with high, medium, and low activations for each superordinate. A variety of feature types (Wu & Barsalou, in press) were used, such as external and internal surface features and components, functions (nonliving things only), entity behaviors (living things only), locations, systemic features, and what things are made of (nonliving things only). The intent was to force participants to consider superordinate concepts generally and not adopt a strategy of focusing on one feature type.

An additional 54 unrelated superordinate trials were constructed, using the same 18 superordinates. Thus, each superordinate was presented three times with a related feature and three times with an unrelated feature so that a superordinate did not cue the response. Related and unrelated features were matched for feature type to prevent participants from using it as a cue (e.g., there were 12 related and 12 unrelated functional features). In addition, the features for approximately three quarters of unrelated trials were taken from concepts in the same domain as the superordinate (i.e., living or nonliving; *bird* <spins webs>), and one quarter from the opposite domain (*bird* <produces radiation>). An additional 30

related and 30 unrelated superordinate feature pairs were included as filler trials. Overall, 50% of the trials corresponded to ''yes'' responses and 50% to ''no.''

Practice items consisted of 10 ''yes'' and 10 ''no'' trials and were comprised of roughly the same proportions of each feature type used for the experimental trials. The categories for the practice trials were *amphibian*, *beverage*, *cleanser*, *dinosaur*, *fashion accessory*, *food*, *jewelry*, *musical instrument*, *stationary*, and *toy*.

5.1.1.3. Procedure: Participants were tested individually on a Macintosh computer using PsyScope (Cohen et al., 1993). Instructions were given verbally and appeared on screen. Participants responded by pressing one of two buttons on a CMU button box, providing accuracy to the nearest millisecond. For each trial, an asterisk appeared for 250 ms, followed by 250 ms of blank screen. A superordinate name then was presented in the center of the screen for 400 ms, followed immediately by a feature name one line below. Both remained on screen until the participant responded. The inter-trial interval was 1500 ms and trials occurred in random order.

Participants were instructed to press the ''yes'' button (using the index finger of their dominant hand) if the feature was characteristic of the category, such that many members of the category can be considered to have the feature (otherwise press the ''no'' button). The criterion of ''many'' was used because it is almost always the case with superordinate concepts that a feature does not apply to every exemplar of the category (Rosch & Mervis, 1975). The task took approximately 20 minutes.

5.1.2. Results

All trials for which an error occurred were removed from decision latency analyses. Decision latencies longer than three standard deviations above the mean of all experimental trials were replaced by that cutoff value (1.7% of the data). The mean decision latency for ''yes'' trials was 824 ms ($SE = 16$ ms) and for ''no'' trials was 901 ms ($SE = 17$ ms). The mean error rate for ''yes'' trials was 11% ($SE = 2\%$) and for ''no'' trials was 5% ($SE = 1\%$).

5.2. Simulation 2

The assumption underlying Simulation 2 is that human feature verification latency is monotonically related to the degree to which a feature is activated during the computation of a superordinate concept. For example, as the representation for *tool* is computed, the activation of <has a handle> should predict verification latency for *tool-has a handle*.

5.2.1. Method
5.2.1.1. Materials: The items were the related target trials used in Experiment 2.

5.2.1.2. Procedure: A feature verification trial was simulated by initializing all feature units between .15 and .25, and presenting a superordinate's wordform for 20 time ticks. Target feature activation was recorded at each time tick.

5.2.2. Results

Target feature activation was used to predict behavioral feature verification latencies after feature length in characters (including spaces), feature length in words, feature length in syllables, and feature frequency (ln[BNC] of all content words in the feature) had been forced into the regression equation (and thus had been partialed out). These lexical variables were partialed out because they are known to influence reading times, but they play no role in the current model. The model significantly predicted target feature verification latency from ticks 4 to 20, with partial correlations ranging from −.28 to −.43, peaking at ticks 9 to 12 (see Table 4). Significant predictions were not expected at the earliest time ticks because the semantic units were initialized to random values, and the network was just beginning to settle.

Somewhat surprisingly, verification latency was not significantly predicted by any of the feature name reading time variables: number of characters, partial $r = -.11$, $p > .4$, number of words, partial $r = .09$, $p > .5$, number of syllables, partial $r = .04$, $p > .7$, or frequency, partial $r = .14$, $p > .3$. Therefore, to confirm that partialing out these variables did not cause spurious results, zero-order correlations were calculated. Correlations were significant for time ticks 3 to 20, with correlations ranging from −.28 to −.43, peaking at ticks 9 and 10.

A potential concern is that activation in the model at different time ticks is correlated, and thus the alpha level may be inflated across the 20 regressions. However, the purpose of

Table 4
Predicting feature verification latencies (Experiment 2) using feature unit activations (Simulation 2)

Time Tick	Partial Correlation
1	.27
2	−.23
3	−.27
4	−.28[*]
5	−.29[*]
6	−.31[*]
7	−.35[*]
8	−.40[**]
9	−.43[**]
10	−.43[**]
11	−.43[**]
12	−.43[**]
13	−.42[**]
14	−.41[**]
15	−.41[**]
16	−.40[**]
17	−.40[**]
18	−.40[**]
19	−.40[**]
20	−.40[**]

Note: *$p < .05$, **$p < .01$.

performing a regression analysis at each tick is to show that the model's ability to predict human performance is not limited to a small window of its temporal dynamics. In fact, predictions were successful for 17 or 18 of 20 ticks, demonstrating that they were robust.

5.3. Discussion

The degree to which superordinate labels activate features in the model successfully predicted feature verification latencies. Experiment 2 and Simulation 2 were identical in methodology to previous feature verification experiments and simulations using basic-level concepts (McRae et al., 1999). This provides further support for the idea that the representations of superordinate concepts are not qualitatively different from basic-level concepts. Just as the learning and representation of these types of concepts are treated the same, so is the computation of semantic features. We return to this point in the General Discussion.

The results of Experiment 2 may also be accounted for by spreading activation theory (Collins & Loftus, 1975). In this theory, stronger criterialities should lead to shorter feature verification latencies. Criterialities are assumed to be directly related to the frequency with which a concept label is paired with the semantic feature. To test this hypothesis, the concept-feature pairs used in Experiment 2 and Simulation 2 were each given a score denoting the proportion of exemplars within the category that possess the relevant feature (like a standardized featural family resemblance measure). These scores predicted verification latencies, partial $r = -.43$, $p < .01$ (with the lexical variables partialed out), suggesting that the criterialities of spreading activation networks would predict verification latencies for superordinate concepts.

Up to this point, our experiments have examined the offline computation of the similarity among exemplar and superordinate concepts (via typicality ratings), and the activation levels of superordinate features. One criticism of Simulations 1 and 2 might be that the attractor network has made predictions that would also be made by family resemblance or spreading activation theory. However, one advantage of our model is that it embodies temporal dynamics. Therefore, Experiment 3 and Simulation 3 demonstrate one interesting way in which in these dynamics are necessary for understanding behavioral phenomena.

6. Experiment 3 and Simulation 3: Superordinate semantic priming

Semantic priming often has been used to gain insight into the structure of semantic memory. In a standard short stimulus onset asynchrony (SOA) semantic priming task, a prime word is presented for a short period of time such as 250 ms, then a target word is presented and the participant responds to it (e.g., by indicating whether it refers to a concrete object, or whether it is indeed a word). If a target word such as *hawk* is preceded by a prime such as *eagle* that is highly similar in terms of featural overlap, decision latency is shorter relative to an unrelated prime, such as *jeep* (Frenck-Mestre & Bueno, 1999; McRae & Boisvert, 1998).

Semantic priming is therefore useful for investigating superordinate–exemplar relations, as in *vegetable* priming *carrot* or *pumpkin*, compared to an unrelated superordinate prime such as *vehicle*. One interesting aspect of priming is that facilitation is sensitive to the relation between concepts, but participants are not asked to explicitly judge those relations. Previous research demonstrates that a superordinate prime facilitates the processing of an exemplar target (Neely, 1991). This result is predicted by all current theories of semantic memory.

An interesting extension is to test whether the magnitude of priming increases with exemplar typicality. In spreading activation theory, *vegetable* should prime a highly typical exemplar such as *carrot* to a greater degree than it primes a less typical exemplar such as *pumpkin*. This prediction obtains because the accessibility of an exemplar given a superordinate prime is determined by the criteriality of the superordinate–exemplar connection weight, and connections to shared features (Collins & Loftus, 1975). A high typicality exemplar has a strong direct link to its superordinate in addition to multiple shared feature nodes. A low typicality exemplar has a weaker direct link to its superordinate, in addition to fewer links to shared features. Therefore, spreading activation theory predicts that the magnitude of priming (relative to an unrelated superordinate prime) increases with typicality.

Distributed feature-based models appear, on the surface, to make the same prediction. An exemplar is processed more quickly if it shares many features with the superordinate than if it shares few features because those features are preactivated. As demonstrated in Experiment and Simulation 1, the degree of featural similarity in the present model predicts typicality ratings. Therefore, it appears that our model would predict that the magnitude of superordinate–exemplar priming increases with typicality. However, as becomes clear below, the behavior of dynamical systems over time, such as in recurrent neural networks, is not always entirely obvious.

Although superordinate priming has been studied in a number of experiments, most have been concerned with expectancy generation (i.e., the ability to anticipate upcoming stimuli) at long SOAs (Keefe & Neely, 1990; Neely, Keefe, & Ross, 1989). However, in one study, Schwanenflugel and Rey (1986) manipulated typicality and used a short 300 ms SOA. They found priming effects for low, medium, and high typicality exemplars, but surprisingly, no interaction between typicality and relatedness. Priming was greatest for medium (73 ms), followed by high (52 ms) and low typicality (30 ms). Note that "relatedness" is used to refer to related versus unrelated control primes, as is customary in priming studies. Thus, for example, a superordinate name would be a related prime for its high, medium, and low typicality exemplars (all three are related to the superordinate because they belong to that category). The unrelated control prime would be the name of another superordinate category, such as *vehicle* for *carrot*.

Schwanenflugel and Rey's (1986) results are surprising. First, they appear to directly contradict all current theories of semantic memory. Second, they conflict with results from analogous basic-level priming experiments. A number of studies found null effects of short SOA priming using prime–target pairs that were thought to be semantically similar (Lupker, 1984; Moss, Ostrin, Tyler, & Marslen-Wilson, 1995; Shelton & Martin, 1992; but see also Frenck-Mestre & Bueno, 1999). However, using significantly more similar pairs

(as determined by similarity ratings), McRae and Boisvert (1998) showed that a basic-level concept was primed to a much greater degree by another basic-level concept that was highly similar than by one that was less similar. That is, *eagle* primes *hawk* to a much greater degree than does *robin*. In fact, at a short 250 ms SOA, both semantic and lexical decisions to a target were facilitated only if the prime was highly similar. At a long 750 ms SOA, priming was significant for both highly and less similar prime-target pairs, but was almost twice as large for the high similarity items (and thus similarity interacted with relatedness). These results appear to be inconsistent with those of Schwanenflugel and Rey (1986) in which superordinate–exemplar priming was not systematically influenced by prime-target similarity (as indexed by typicality).

McRae and Boisvert (1998) priming effects were simulated using a feature-based connectionist attractor network similar to the present model. Cree, McRae, and McNorgan (1999) simulated semantic priming by assuming that with a short SOA, a prime is partially activated prior to the presentation of the target word. Thus, the prime was presented to the network for 15 time ticks (it was trained using 20 ticks), and then the target was presented for an additional 20 ticks. Cree et al. found larger priming effects for highly than for less similar prime-target pairs, consistent with McRae and Boisvert's experiments. This further supports the notion that the magnitude of priming increases as similarity between prime and target increases, and again, is contrary to Schwanenflugel and Rey (1986).

In summary, previous theoretical and empirical accounts suggest that Schwanenflugel and Rey's (1986) results are implausible. Therefore, Experiment 3 was a replication of their experiment with two differences. First, a new and larger set of items were derived from our norms. Second, a two (related vs. unrelated) by two (high vs. low typicality) design was adopted by removing Schwanenflugel and Rey's medium typicality condition. Because replicating their results entails a null relatedness by typicality interaction, the two by two design promoted finding an interaction if there was one to be found. Experiment 3 was then simulated using the same technique as Cree et al. (1999). To foreshadow the results, Schwanenflugel and Rey's effect was replicated in both Experiment 3 and Simulation 3.

6.1. Experiment 3

6.1.1. Method
6.1.1.1. Participants: Fifty-three undergraduates at the University of Western Ontario participated in course credit, 25 in List 1, and 28 in List 2. One participant from List 2 was dropped because his/her error rate was an extreme outlier, and two were participants from List 2 were dropped because their decision latencies were extreme outliers.

6.1.1.2. Materials: Fourteen superordinate categories served as primes—*appliance*, *bird*, *carnivore*, *clothing*, *container*, *fruit*, *furniture*, *insect*, *mammal*, *tool*, *utensil*, *vegetable*, *vehicle*, *weapon*—and two exemplars from each category served as targets, one low in typicality and one high (see Appendix C). The distributions of typicality ratings were non-overlapping: low typicality ($M = 5.86$, $SE = .22$), high typicality ($M = 8.06$, $SE = .16$), $t(26) = 8.13$, $p < .001$. The groups also differed in mean superordinate–exemplar cosine

from the model; with the exception of one item, these distributions were nonoverlapping, low typicality (*M* = .46, *SE* = .01), high typicality (*M* = .57, *SE* = .01), *t*(26) = 6.52, *p* < .001.

Two lists were constructed so that no participant was presented with a target more than once or a prime more than twice. For each list, half of the targets were related (a member of the superordinate category), and half were unrelated. Each half was split equally between high and low typicality targets. The unrelated primes were created by re-pairing related primes and targets.

The superordinate primes were the same for both typicality groups. Because the targets differed, they were equated on a number of variables known to influence word reading, and thus the potential for priming effects. These are presented in Table 5. The two typicality groups were equated on word length (number of characters and syllables), printed word frequency (ln(freq) from the BNC), rated concept familiarity, mean number of features per concept, and Coltheart *N* (a measure of orthographic neighborhood density). This equating process, and differentiating the groups on typicality ratings, necessitated reducing the number of categories from 20 to 14.

To minimize participant expectancies, the same 84 fillers were added to each list. All filler trials consisted of a superordinate followed by a basic-level concept. As in experimental trials, all filler primes were presented twice and all targets once. Because a concreteness decision task was used, 28 fillers were unrelated concrete–abstract pairs (*spice–notion*, requiring a ''no'' response), 28 were unrelated abstract–concrete pairs (*religion–razor*, requiring a ''yes'' response), 14 were unrelated abstract–abstract pairs (*emotion–strategy*, ''no''), and 14 were related abstract–abstract pairs (*crime–fraud*, ''no''). Thus, half of the trials required ''yes'' responses, half of the primes were concrete, and the relatedness proportion was .25. In addition, 25% of the trials were concrete–concrete pairs, 25% were concrete–abstract, 25% were abstract–abstract, and 25% were abstract–concrete.

There were 22 practice items, of which six were related (27%), and 11 had targets denoting concrete objects (50%). Six pairs were concrete–abstract (27%), six were concrete–concrete (27%), five were abstract–concrete (23%), and five were abstract–abstract (23%). Thus, proportions were similar to the experimental trials. There were 11 superordinates used

Table 5
Equated variables for high- and low-typicality exemplars in Experiment 3 (semantic priming)

| Factor | Typicality | | *F*(1, 26) | *p* |
	High	Low		
Length in characters	6.00	5.93	.02	>.9
Length in syllables	1.79	1.79	.00	1.0
Frequency: ln(BNC)	6.71	6.74	.01	>.9
Familiarity	6.47	6.35	.03	>.8
Number of features/concept	13.00	13.14	.02	.9
Coltheart *N*	3.14	3.21	.00	>.9

Note: BNC, British National Corpus.

as practice primes so that each prime appeared twice and each target once. No experimental primes or targets appeared in the practice trials.

6.1.1.3. Procedure: Participants performed the experiment individually. The apparatus was the same as in Experiment 2. A trial consisted of an asterisk for 250 ms, followed by 250 ms of blank screen, the prime for 200 ms, a 50 ms blank screen inter-stimulus interval, and then the target. The target remained on screen until the participant responded, and the inter-trial interval was 1500 ms. Participants were instructed to read and pay attention to the first word, but not to respond to it. They were instructed to press the ''yes'' button using the index finger of their dominant hand if the second word referred to a concrete object, which was defined as ''something that is touchable.'' The task took approximately 15–20 min.

6.1.1.4. Design: The independent variables were typicality (high vs. low) and relatedness (related vs. unrelated). List was included as a between-participants dummy variable and item rotation group as a between-items dummy variable to stabilize variance that may result from rotating participants and items over lists (Pollatsek & Well, 1995). Relatedness was within-participants (F_1) and within-items (F_2), whereas typicality was within-participants but between-items. The dependent measures were decision latency and the square root of the number of errors (Myers, 1979).

6.1.2. Results
6.1.2.1. Decision latency: Mean decision latencies for each condition are presented in Table 6. All trials for which an error occurred were removed from the decision latency analyses. Trials longer than three standard deviations above the mean of the experimental trials were replaced by that value (2% of the data).

The results replicated Schwanenflugel and Rey (1986). Relatedness did not interact with typicality, both Fs < 1. In fact, the priming effect for low-typicality pairs (34 ms) was slightly larger than for high-typicality pairs (27 ms). Decision latencies for related trials ($M = 627$ ms, $SE = 18$ ms) were significantly shorter than for unrelated trials ($M = 658$ ms, $SE = 18$ ms), $F_1(1, 48) = 12.71$, $p = .001$, $F_2(1, 24) = 9.52$, $p < .01$. There was no main effect of typicality, both Fs < 1.

Table 6
Decision latencies (ms) for Experiment 3 (semantic priming)

| | Typicality | | | |
| | High | | Low | |
	M	*SE*	*M*	*SE*
Unrelated	654	19	661	19
Related	627	19	627	19
Priming effect	27		34	

6.1.2.2. Error rates: The mean error rates were as follows: related high typicality, 3% (SE = 1%); unrelated high typicality, 5% (SE = 1%); related low typicality, 5% (SE = 1%); and unrelated low typicality, 3% (SE = 1%). There was no main effect of relatedness, both Fs < 1, no main effect of typicality, both Fs < 1, but a marginally significant interaction, $F_2(1, 48) = 3.91$, $p < .1$, $F_2(1, 24) = 3.99$, $p < .1$. However, the error rates were extremely low, thus exaggerating the influence of a single error from a single participant.

6.2. Simulation 3

Superordinate–exemplar priming was simulated using the prime-target pairs from Experiment 3. The methodology was the same as in Cree et al. (1999) and Masson (1995).

6.2.1. Method
6.2.1.1. Procedure: Prior to presenting the superordinate prime, the activations of all semantic units were initialized to random values between .15 and .25 (as with training and Simulations 1 and 2). The superordinate's wordform was input for 15 ticks. Then, with the network in the state representing the superordinate, the target's wordform was input. The network settled to the representation of the target for the remaining 20 ticks. CEE relative to the target, with an error radius of .2 (because features with activations >.8 were deemed correctly on during training), was then computed at the semantic layer over the entire 35 ticks for each trial. Note that this means that, for the time ticks during which the superordinate prime was presented (*vegetable*), we measured error relative to the basic-level target for that trial (e.g., *spinach*).

6.2.1.2. Design: Two analyses of variance were conducted. The first was performed on the first 16 ticks (tick −15 to tick 0 relative to target onset). The second analysis concerned the final 19 ticks (ticks 1 to 19 relative to target onset) during which the basic-level target was presented to the model. Although the target was presented at tick 0, its influence is not observed at the semantic layer until tick 1 when activation spreads from the input to the output layer, and therefore the output of the model at tick 0 is considered with the prime for the analyses. In both analyses, the independent variables were typicality (low and high), relatedness (related vs. unrelated), and time tick. Item rotation group was included as a (between-items) variable to account for changes in variance associated with using different prime-target pairings across the two lists (Pollatsek & Well, 1995). Relatedness and time tick were within-items and typicality was between-items. The dependent measure was CEE.

6.2.2. Results
The settling profile for each condition is presented in Fig. 2. While the model processed the prime, the differences among conditions reflect manipulations of relatedness and typicality. Recall that cross entropy is measured relative to the target (i.e., the difference between what is being computed from the superordinate prime and the representation of the ensuing target). As expected, when an unrelated prime such as *clothing* is computed when the target is a vegetable, the error trajectory is roughly equivalent whether the target is high (*spinach*)
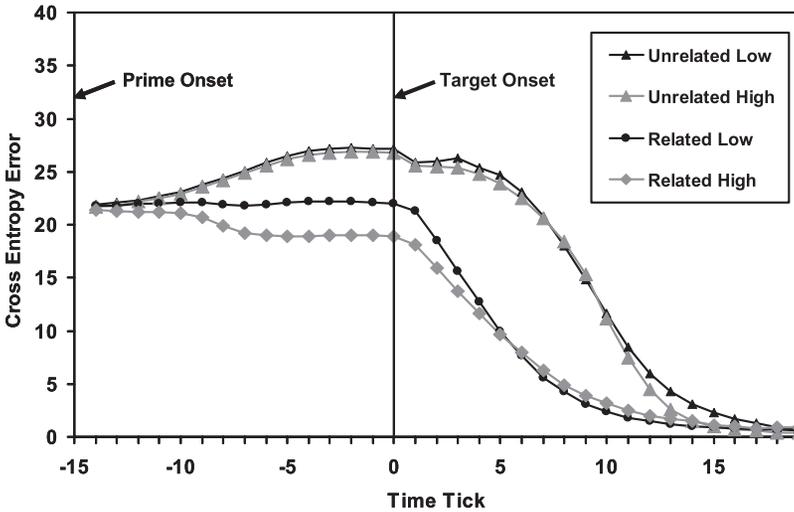
Fig. 2. Mean cross entropy error relative to the target concept over time in Simulation 3 (semantic priming).

or low typicality (*mushroom*), because neither resembles *clothing*. Cross entropy increases over time in these conditions because the features of the prime that become activated are inconsistent with the unrelated targets. For the related conditions, as the model settles into a prime representation (*vegetable*), CEE relative to a high-typicality target (*spinach*) is lower than to a low-typicality target (*mushroom*), particularly at later ticks (closer to target onset).

After target onset, consistent with the unrelated conditions prior to target onset, the unrelated high- and low-typicality items demonstrate similar settling patterns. However, in the related conditions, the differences in cross entropy between the high- and low-typicality conditions abate, and the two lines converge after a few ticks.

6.2.2.1. Prime analyses: The ANOVA on the first 16 ticks (tick −15 to tick 0 relative to target onset), during the presentation of the superordinate, revealed a main effect of time, $F(15, 360) = 1,085,396.00$, $p < .001$, and no main effect of typicality, $F < 1$. However, overall, unrelated primes ($M = 61.66$, $SE = .98$) yielded greater error relative to the target exemplars than did related primes ($M = 58.10$, $SE = .91$), $F(1, 24) = 216.73$, $p < .001$. This difference is greater for high-typicality than for low-typicality pairs, particularly at later ticks, resulting in a relatedness by typicality interaction, $F(1, 24) = 12.42$, $p < .01$. Relatedness and typicality further interacted with time because semantic activation was initially random, and at early ticks the model had not sufficiently computed its representation, $F(15, 360) = 4.62$, $p < .001$. Therefore, differences among the four conditions were not apparent until the model had begun to settle into a superordinate representation. For the same reason, time also interacted with relatedness, $F(15, 360) = 101.43$, $p < .001$, and typicality, $F(15, 360) = 5.31$, $p < .001$. This pattern of error trajectories was expected given that the cosines between the primes and targets for the high- and low-typicality conditions were from nonoverlapping distributions. In addition, the difference between the related high- and low-typicality conditions was expected given that the model predicted offline typicality ratings in Simulation 1.

6.2.2.2. Target analyses: Consistent with Experiment 3 and Schwanenflugel and Rey (1986), in an ANOVA on ticks 16 to 35 (ticks 1 to 19 relative to target onset), when activation of the target had reached the semantic layer, typicality and relatedness did not interact, $F < 1$. CEE was lower for related ($M = 5.73$, $SE = .47$) than for unrelated prime-target pairs ($M = 12.57$, $SE = .71$), $F(1, 24) = 134.05$, $p < .001$, showing an overall priming effect. There was no main effect of typicality, $F < 1$. A main effect of time obtained in that error gradually decreased as the model settled into the target basic-level concept, $F(18, 432) = 328.31$, $p < .001$. Relatedness and time also interacted because differences in cross entropy between related and unrelated conditions disappeared over the last few ticks as the concepts settled fully, $F(18, 432) = 93.23$, $p < .001$. Planned comparisons were conducted at each tick to further investigate the relatedness by time interaction at each level of typicality. For low-typicality items, there was a significant difference between the related and unrelated conditions for ticks 2–11 (relative to target onset), and similarly, for the high-typicality items, differences between related and unrelated conditions were significant for ticks 1–11.

6.3. Discussion

Simulation 3 is consistent with Experiment 3. Targets preceded by a related superordinate were facilitated for both the model and humans. In both humans and the model, the magnitude of priming did not vary systematically with the typicality of the exemplar targets; there was no hint of an interaction between relatedness and typicality for either the model or humans, with Experiment 3 and Simulation 3 replicating Schwanenflugel and Rey (1986). In fact, the magnitude of priming was slightly larger for low-typicality items than for high-typicality items in Experiment 3.

Note that these results do not mean that the model would be incapable of simulating category verification latencies. In category verification, participants are presented with items such as ''*vegetable carrot*,'' and asked to indicate whether a carrot is a member of the vegetable category. In Simulation 3, we predicted concreteness decision latencies using CEE as the target settled. If our model was to be used to predict category verification latencies, the similarity between the superordinate and basic-level exemplar concepts would be key, not the settling pattern of the exemplar. These similarities are sensitive to typicality, as shown in Simulation 1 and by the Simulation 3 typicality by relatedness interaction that occurred prior to target onset. Therefore, the intriguing result is that this difference in baseline similarity did not influence the magnitude of priming.

The results of Experiment 3 and Schwanenflugel and Rey (1986) are problematic for Collins and Loftus' (1975) spreading activation theory. In that model, the same static similarity measure is used to account for typicality ratings, superordinate–exemplar priming, and priming between basic-level concepts.

These results also appear, on the face of it, to contradict predictions of a feature-based distributed network. However, Simulation 3 produced priming effects of similar magnitude for both typicality levels, even though Simulation 1 predicted typicality ratings. Also recall that a model similar in architecture to the present one produced differential priming effects between basic-level concepts of varying similarity (Cree et al., 1999), and in fact, those

simulations were successfully replicated using the present model (although we do not present the simulations herein). The model's ability to account for these seemingly inconsistent results depends on two factors: the nature of superordinate versus basic-level computed representations, and the network's temporal dynamics.

Superordinate concepts are represented quantitatively differently than are basic-level concepts in the model. After settling, most features of superordinate concepts are activated around .5, whereas features of basic-level concepts are activated close to one. Offline typicality ratings are not influenced by these differences because they are not sensitive to computational temporal dynamics. Typicality ratings were simulated by computing each representation separately and then comparing them, as humans presumably do. In contrast, the priming simulation illustrates one consequence of these representational differences.

A feature's activation in the network is influenced by both the total input to the unit from all incoming connections, and the sigmoidal activation function (see Fig. 3), which transforms net input into an activation value between zero and one. As a result, features activated close to .5 are relatively easy to turn on or off. This occurs because the slope of the sigmoidal activation function close to the midpoint is relatively steep. In contrast, it is more difficult (i.e., requires larger changes in net input) to change the activation of units that have extremely high or low net input and thus are activated at essentially one or zero. These exemplar features are illustrated on the extremes of Fig. 3.

When a related superordinate prime (*vegetable*) is computed, followed by an exemplar target (*spinach*), a feature that is shared by the prime and target (e.g., <is green>) contributes to facilitation because it is preactivated by the prime. If a feature is activated by the superordinate that is not shared by the exemplar (<is green> for *pumpkin*), this feature is relatively easy to turn off while *pumpkin* is being computed because it lies on the steep part of the activation function. Because at least some features of a superordinate are shared by a related exemplar, there is indeed a priming effect relative to an unrelated superordinate. However, because features activated by the superordinate name that are not part of the exemplar's representation are relatively easily turned off, there is not a great deal of
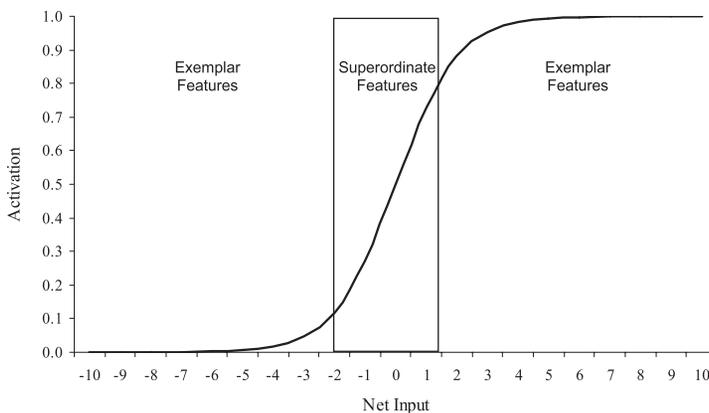

Fig. 3. Sigmoidal activation function and activation levels of superordinate and basic-level exemplar features.

inhibition due to differing features. Thus, the network produces equivalent priming across a range of typicality.

Contrast this with priming between two basic-level concepts in which facilitation is highly sensitive to similarity (McRae & Boivert, 1998). The prime concept consists of features that are essentially turned fully on or off after settling. Therefore, features of the prime that are not possessed by the target slow activation of the target concept because they are difficult to de-activate. Even if the prime and target share a moderate number of features, facilitation is dampened because of inhibition due to mismatching features. In contrast, if there is a high degree of featural overlap between prime and target, then there are few prime features that need to be de-activated while computing the target concept. Thus, a substantial influence of the degree of similarity is obtained for pairs of basic-level concepts in both humans and the model.

In summary, due to the relative difficulty of changing the activations of units that are activated virtually to one or zero (as in basic-level primes) versus the relative ease of changing activations that are on the steep section of the sigmoidal function (as in superordinate primes), the influence of prime-target similarity is much more pronounced in basic-level to basic-level priming. This aspect of the temporal dynamics of the computation of word meaning, in combination with quantitative differences between computed superordinate versus basic-level concepts, explains the results regarding the seemingly contradictory influences of similarity. Thus, offline measures of similarity do not provide the entire picture. Understanding the richness of the relevant results requires consideration of the temporal dynamics of similarity.

7. General discussion

Our experiments and simulations demonstrate that a number of interesting empirical semantic memory phenomena can be simulated using a model with a single representational layer, that is, without an implemented hierarchical structure. The model learned basic-level and superordinate concepts by treating both types similarly, as distributed sets of features in a flat attractor network. It computed reasonable superordinate representations and distinguished between category members versus nonmembers. In Simulation 1 and Experiment 1, the model accounted for an offline measure of the graded structure of categories, typicality ratings. In Simulation 2 and Experiment 2, the activations of superordinate features predicted feature verification latencies. Simulation 3 provided novel insight into counter-intuitive results regarding superordinate semantic priming by combining quantitative differences between superordinate and basic-level featural activations with the computational dynamics of similarity.

Thus, our model is an advance over past models of semantic memory in a few important ways. For example, neither family resemblance, considered the gold standard for assessing the structure of natural categories for decades, nor explicit superordinate representations, had been implemented in a learning system. Our model learns superordinate representations that have family resemblance-like structure, and it performs comparably to family resemblance when assessing typicality. Possibly of greatest import, our model allowed for these

representational issues to be studied in conjunction with processing dynamics, and thus the computation of superordinates over time.

7.1. Commonalities and differences between basic-level and superordinate concepts

A central theoretical assumption underlying our modeling is that all learned concepts, regardless of their putative ''level,'' are treated similarly in many respects. Every concept was learned by presenting the model with the name for a class of entities or objects, and pairing it with an instance of that category on each learning trial, just as a human would have numerous labeled perceptual experiences. In this sense, our training regime is a direct statement about how people learn. For superordinates such as *fruit* and *furniture*, people do not develop representations by being explicitly taught that ''an apple is a fruit''; rather, incidental, implicit learning is key. Labeling plays a critical role in this process. As Brown (1958) aptly states, ''the child's vocabulary is more immediately determined by the naming practices of adults'' (p. 18). Furthermore, as has been demonstrated in recent investigations of conceptual development (Plunkett et al., 2007), labels are useful pointers to concepts only when they cohere with semantic structure already present in the environment. Thus, despite equivalence in the principles underlying the learning and computation of both superordinate and basic-level concepts (and all those in between), the model and humans arrive at representations for these concepts that differ in important ways.

One such way is that the patterns of feature activations in the model systematically vary, and we believe this has a direct correspondence in humans. There are clear representational differences between (many of the) superordinate concepts, such as *furniture*, and the basic-level concepts, such as *chair*, whereby the degree of featural overlap of the exemplars denoted by the category names differs markedly. For example, there are considerably more shared features (and stronger correlations among them) among *armchairs*, *rocking chairs*, and *kitchen chairs* that comprise the category *chair*, than among *tables*, *beds*, and *cabinets* that comprise *furniture*. Of course, specific instances of a subordinate category such as *rocking chair* are even more similar to one another. That is, some words are used consistently to point out bundles of features that, taken together, form relatively concrete and imageable object representations. Some other words, paired with a more varied set of perceptual instances, can be used to highlight important subsets of features that appear consistently together across objects. Therefore, the semantic system must ascertain which subsets are being referred to by the label. The result is that for many superordinate labels, the features are more varied in activation state than are the features activated by a basic-level label. One consequence of this is that these superordinates are considerably less imageable and thus might be considered as somewhat abstract.

An extension of this view is that the model naturally learns concepts for which their level in a purported hierarchy is unclear. For example, the model developed a representation for *bird* that fell between superordinates like *furniture* and basic levels like *chair*, in that *bird* had numerous features that were, more or less, fully activated, as well as some with intermediate activations. This occurred even though the training procedure for *bird* was identical to all other superordinates. Interestingly, *bird* is one concept that has sometimes been identified as a basic-level concept, and at other times as a superordinate (Rosch et al., 1976). This

underscores Randall's (1976) argument that people do not care about the ''level'' in the naming hierarchy at which a label might be described as residing.

Obviating the need for a strict taxonomy also removes any potential problems with the facts that many concepts do not belong to any category (such as *ashtray*), whereas numerous others are exemplars of multiple categories, such as *knife*. During learning of any type of object concept, a perceived entity (or set of entities) is present, echoing Randall's (1976) assertion that the physical features of an object play a key role in learning, regardless of how they do, or do not, fit into any purported hierarchy.

Differences in activation levels also provided insight into superordinate to basic-level priming. Along with the model's temporal dynamics of its computations, these differences enabled understanding the seemingly inconsistent results regarding superordinate to basic-level priming, typicality ratings, and priming between two basic-level concepts in both the model and humans.

7.2. Neural representation

We believe that our model is more consistent with what is known about the neural representation of concepts than is the idea of hierarchical representation. It is widely accepted that knowledge of object concepts is represented across multiple modality-specific brain regions (Goldberg, Perfetti, & Schneider, 2006; Martin, 2007). For example, visual information is stored in regions that are distinct from those that store information about taste, sound, or an object's function. Modality-specific regions were not implemented in our model, although there is no reason why they could not be (see Cree & McRae's, 2003, brain region feature classification scheme). If they were, then one possibility is that superordinate concepts are distributed across precisely the same regions as basic-level concepts, in the way that they share representational units in the model. The main difference would be the level of activation of the assemblies of neurons involved, resulting from the degree to which regularities exist for information within any modality. Although not a driving force for our research, this type of model maximizes cognitive economy.

On the other hand, it could be that the anterior temporal pole, an area argued to be a high-level convergence zone, might function much like the superordinate levels of a hierarchical representational system, with lower level knowledge being distributed across modality-specific pathways (Patterson, Nestor, & Rogers, 2007). Note, however, that this type of hierarchy would fundamentally differ from that of Collins and Quillian (1969), in that all feature-based knowledge would be stored at the lower levels, and a strict taxonomic hierarchy would not be transparently implemented.

7.3. General features

One argument against using feature norms is that features common to a large number of concepts often are not listed, even though participants are aware of them. Participants seldom list features such as a *bear* <has a heart> and <is alive>, or that a *car* <is built in a factory> and <is owned by people>. Instead, they are biased toward producing distinguishing

rather than shared features because distinguishing information disambiguates a concept from other similar concepts (McRae et al., 2005). Thus, our model did not contain those general features.

One issue, therefore, is whether their omission influenced our results in a systematic manner. It seems that a lack of general features should have little influence or might be detrimental. For example, adding <has veins> and <has eyes> to all animals should not change the predictability of graded structure. It would presumably increase the similarity of all animal exemplars to their superordinate, but the relative differences in superordinate–exemplar similarity would remain. Another way to think about this is that such general features may play little role in tasks such as typicality rating because all category members share them. The same argument can be made for the priming simulations. If all exemplars of a category include additional shared features with their superordinate prime, it should facilitate the processing of low- and high-typicality exemplars equally.

7.4. Model architecture

One aspect of the model's architecture that was somewhat atypical is that it did not include a hidden layer in which patterns of activation are taken as abstracted representations (Plaut, 2002; Rumelhart & Todd, 1993). We avoided using a hidden layer to obviate any possible argument that our model might actually be encoding different ''levels'' of abstraction in those units. We suspect that using a hidden layer would not qualitatively change the results of our simulations because semantic regularities and the influence of label-concept pairings would continue to determine what the network learns.

7.5. Conclusions

Our research demonstrates that it is not necessary to explicitly implement a hierarchy to produce behavior that appears to be hierarchically driven. Thus, the idea of a hierarchy might be best thought of as descriptive of behavior, rather than being literally applicable to mental representation (i.e., an emergent phenomenon). Our approach also avoids issues that arise when researchers attempt to apply a strict hierarchy on sets of concepts that actually do not form such a hierarchy. Most pertinent to this special issue, we used an attractor network that enabled studying the temporal dynamics of the computation of various types of concepts. This led to Jeff Elman's suggestion to highlight the theoretical import of what he coined ''the dynamics of similarity.''

Notes

1. Other similar proportions (e.g., 75% basic-level concepts, 25% superordinates) were tested and there were no notable changes in superordinate or basic-level representations.
2. Spearman's nonparametric rank-order correlations were also computed as a reliability check because it is less sensitive than Pearson's product-moment correlations to

outliers. The only notable difference was that Spearman's correlation between cosine and typicality was nonsignificant for *musical instruments*. Furthermore, visual inspection of the 20 scatterplots revealed two obvious outliers (*lamp* in *furniture,* and *porcupine* in *carnivore*). With these points removed, the Pearson correlation for *furniture* was .45, $p = .052$, and for *carnivore* was .51, $p < .02$.

Acknowledgments

References

Anderson, J. R. (1983). A spreading activation theory of memory. *Journal of Verbal Learning and Verbal Behavior*, *22*, 261–295.

Barsalou, L. W. (2003). Abstraction in perceptual symbol systems. *Philosophical Transactions of the Royal Society of London: Series B*, *358*, 1177–1187.

Berlin, B. (1972). Speculations on the growth of ethnobotanical nomenclature. *Journal of Language & Society*, *1*, 63–98.

Berlin, B., Breedlove, D. E., & Raven, P. H. (1973). General principles of classification and nomenclature in folk biology. *American Anthropologist*, *75*, 214–242.

Booth, A. E., & Waxman, S. R. (2003). Mapping words to the word in infancy: Infant's expectations for count names and adjectives. *Journal of Cognition and Development*, *4*, 357–381.

Brown, R. (1958). How shall a thing be called? *Psychological Review*, *65*, 14–21.

Cohen, J. D., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: A new graphic interactive environment for designing psychology experiments. *Behavior 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.

Cree, G. S., McNorgan, C., & McRae, K. (2006). Distinctive features hold a privileged status in the computation of word meaning: Implications for theories of semantic memory. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *32*, 643–658.

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

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

Deese, J. (1965). *The structure of associations in language and thought*. Baltimore, MD: Johns Hopkins Press.

Francis, W. N., & Kucera, H. (1982). *Frequency analysis of English usage: Lexicon and grammar*. Boston: Houghton-Mifflin.

Frenck-Mestre, C., & Bueno, S. (1999). Semantic features and semantic categories: Differences in rapid activation of the lexicon. *Brain and Language*, *68*, 199–204.

Fulkerson, A. L., & Haaf, R. A. (2003). The influence of labels, non-labeling sounds, and source of auditory input on 9- and 15-month-olds' object categorization. *Infancy*, *4*, 349–369.

Gelman, R. (1990). First principles organize attention to and learning about relevant data: Number and the animate/inanimate distinction as examples. *Cognitive Science*, *14*, 79–106.

Goldberg, R. F., Perfetti, C. A., & Schneider, W. (2006). Perceptual knowledge retrieval activates sensory brain regions. *The Journal of Neuroscience*, *26*, 4917–4921.

Hinton, G. E. (1981). Implementing semantic networks in parallel hardware. In G. E. Hinton & J. A. Anderson (Eds.), *Parallel models of associative memory* (pp. 161–187). Hillsdale, NJ: Erlbaum.

Hinton, G. E. (1986). Learning distributed representations of concepts. In *Proceedings of the eighth annual conference of the Cognitive Science Society* (pp. 1–12). Hillsdale, NJ: Erlbaum.

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

Hodges, J. R., Graham, N., & Patterson, K. (1995). Charting the progression in semantic dementia: Implications for the organization of semantic memory. *Memory*, *3*, 463–495.

Jolicoeur, P., Gluck, M. A., & Kosslyn, S. M. (1984). Pictures and names: Making the connection. *Cognitive Psychology*, *16*, 243–275.

Keefe, D. E., & Neely, J. H. (1990). Semantic priming in the pronunciation task: The role of prospective prime-generated expectancies. *Memory and Cognition*, *18*, 289–298.

Lucariello, J., & Nelson, K. (1986). Context effects on lexical specificity in maternal and child discourse. *Journal of Child Language*, *13*, 507–522.

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

Macario, J. F. (1991). Young children's use of color in classification: Foods and canonically colored objects. *Cognitive Development*, *6*, 17–46.

Mandler, J. M., Bauer, P., & McDonough, L. (1991). Separating the sheep from the goats: Differentiating global categories. *Cognitive Psychology*, *23*, 263–298.

Markman, A. R., & Stilwell, C. H. (2001). Role-governed categories. *Journal of Experimental and Theoretical Artificial Intelligence*, *13*, 329–358.

Martin, A. (2007). The representation of object concepts in the brain. *Annual Review of Psychology*, *58*, 25–45.

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

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

McClelland, J. L., & Rumelhart, D. E. (1985). Distributed memory and representation of general and specific information. *Journal of Experimental Psychology: General*, *114*, 159–197.

McCloskey, M. E., & Glucksberg, S. (1978). Natural categories: Well defined or fuzzy sets? *Memory & Cognition*, *6*, 462–472.

McRae, K. (2004). Semantic memory: Some insights from feature-based connectionist attractor networks. In B. H. Ross (Ed.), *Psychology of learning and motivation: Advances in research and theory* (Vol. 45, pp. 41–86). San Diego, CA: Academic Press.

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

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

McRae, K., Cree, G. S., Westmacott, R., & de Sa, V. (1999). Further evidence for feature correlations in semantic memory. *Canadian Journal of Experimental Psychology: Special Issue on Models of Word Recognition*, *53*, 360–373.

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.

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.

Murphy, G. L., & Lassaline, M. E. (1997). Hierarchical structure in concepts and the basic level of categorization. In K. Lamberts & D. Shanks (Eds.), *Knowledge, concepts, and categories* (pp. 93–132). Hove, England: Psychology Press.

Murphy, G. L., & Smith, E. E. (1982). Basic level superiority in picture categorization. *Journal of Verbal Learning and Verbal Behavior*, *21*, 1–20.

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

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

Neely, J. H., Keefe, D. E., & Ross, K. L. (1989). Semantic priming in the lexical decision task: Roles of prospective prime-generated expectancies and retrospective semantic matching. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*, 1003–1019.

Patterson, K., Nestor, P. J., & Rogers, T. T. (2007). Where do you know what you know? The representation of semantic knowledge in the human brain. *Nature Reviews Neuroscience*, *8*, 976–988.

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

Pecher, D., Zeelenberg, R., & Barsalou, L. W. (2003). Verifying properties from different modalities for concepts produces switching costs. *Psychological Science*, *14*, 119–124.

Plaut, D. C. (2002). Graded modality-specific specialization in semantics: A computational account of optic aphasia. *Cognitive Neuropsychology*, *19*, 603–639.

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

Plaut, D. C., & Shallice, T. (1994). *Connectionist modeling in cognitive neuropsychology: A case study*. Hillsdale, NJ: Erlbaum.

Plunkett, K., & Elman, J. L. (1997). *Exercises in rethinking innateness: A handbook for connectionist simulations*. Cambridge, MA: MIT Press.

Plunkett, K., Hu, J.-F., & Cohen, L. B. (2007). Labels can override perceptual categories in early infancy. *Cognition*, *106*, 665–681.

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.

Quinn, P. C., & Johnson, M. H. (2000). Global-before-basic object categorization in connectionist networks and 2-month-old infants. *Infancy*, *1*, 31–46.

Randall, R. A. (1976). How tall is a taxonomic tree? Some evidence for dwarfism. *American Ethnologist*, *3*, 543–553.

Rips, L. J., Shoben, E. J., & Smith, E. E. (1973). Semantic distance and the verification of semantic relations. *Journal of Verbal Learning and Verbal Behavior*, *12*, 1–20.

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

Rogers, T. T., & McClelland, J. L.(2005). A parallel distributed processing approach to semantic cognition: Applications to conceptual development. In L. Gershkoff-Stowe & D. H. Rakison (Eds.), *Building object categories in developmental time* (pp. 335–387). Mahwah, NJ: Erlbaum.

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*, *7*, 573–605.

Rumelhart, D. E. (1990). Brain style computation: Learning and generalization. In S. F. Zornetzer, J. L. Davis, & C. Lau (Eds.), *An introduction to neural and electronic networks* (pp. 405–420). San Diego, CA: Academic Press.

Rumelhart, D. E., & Todd, P. M. (1993). Learning and connectionist representation. In D. E. Meyer & S. Kornblum (Eds.), *Attention and performance XIV: Synergies in experimental psychology, artificial intelligence, and cognitive neuroscience* (pp. 3–30). 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.

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

Smith, E. E., Shoben, E. J., & Rips, L. J. (1974). Structure and process in semantic memory: A featural model for semantic decision. *Psychological Review*, *81*, 204–241.

Solomon, K. O., & Barsalou, L. W. (2004). Perceptual simulation in property verification. *Memory & Cognition*, *32*, 244–259.

Vigliocco, G., Vinson, D. P., Lewis, W., & Garrett, M. F. (2004). Representing the meanings of object and action words: The featural and unitary semantic space hypothesis. *Cognitive Psychology*, *48*, 422–488.

Warrington, E. K. (1975). Selective impairment of semantic memory. *Quarterly Journal of Experimental Psychology*, *27*, 635–657.

Waxman, S. R., & Markow, D. B. (1995). Words as invitations to form categories: Evidence from 12- to 13-month old infants. *Cognitive Psychology*, *29*, 257–302.

Wisniewski, E. J., & Murphy, G. L. (1989). Superordinate and basic category names in discourse: A textual analysis. *Discourse Processes*, *12*, 245–261.

Wu, L. L., & Barsalou, L. W. (in press). Perceptual simulation in conceptual combination: Evidence from property generation. *Acta Psychologica*.

Appendix A: Superordinate categories and their basic-level exemplars. Other basic-level concepts were trained in the model but were not part of any superordinate category

Superordinate	Basic-Level Exemplars
animal (133)	alligator, bat (animal), bear, beaver, beetle, bison, blackbird, bluejay, budgie, buffalo, bull, butterfly, buzzard, calf, camel, canary, caribou, cat, caterpillar, catfish, cheetah, chickadee, chicken, chimp, chipmunk, clam, cockroach, cougar, cow, coyote, crab, crocodile, crow, deer, dog, dolphin, donkey, dove, duck, eagle, eel, elephant, elk, emu, falcon, fawn, finch, flamingo, fox, frog, giraffe, goat, goldfish, goose, gopher, gorilla, groundhog, guppy, hamster, hare, hawk, hornet, horse, hyena, iguana, lamb, leopard, lion, lobster, mackerel, mink, minnow, mole (animal), moose, moth, mouse, nightingale, octopus, oriole, ostrich, otter, owl, ox, panther, parakeet, partridge, peacock, pelican, penguin, perch, pheasant, pig, pigeon, platypus, pony, porcupine, pigeon, rabbit, raccoon, rat, rattlesnake, raven, robin, rooster, salamander, salmon, seagull, seal, sheep, shrimp, skunk, snail, sparrow, spider, squid, squirrel, starling, stork, swan, tiger, toad, tortoise, trout, tuna, turkey, turtle, vulture, walrus, wasp, whale, woodpecker, worm, zebra

Appendix A
(Continued)

Superordinate	Basic-Level Exemplars
appliance (14)	blender, dishwasher, drill, fan (appliance), freezer, fridge, kettle, microwave, mixer, oven, radio, stove, telephone, toaster
bird (39)	blackbird, bluejay, budgie, buzzard, canary, chickadee, chicken, crow, dove, duck, eagle, emu, falcon, finch, flamingo, goose, hawk, nightingale, oriole, ostrich, owl, parakeet, partridge, peacock, pelican, penguin, pheasant, pigeon, raven, robin, rooster, seagull, sparrow, starling, stork, swan, turkey, vulture, woodpecker
carnivore (19)	alligator, bear, buzzard, cheetah, cougar, coyote, crocodile, dog, eagle, falcon, fox, hawk, hyena, leopard, lion, panther, porcupine, tiger, vulture
clothing (39)	apron, belt, blouse, boots, bra, camisole, cap (hat), cape, cloak, coat, dress, earmuffs, gloves, gown, hose (leggings), jacket, jeans, leotards, mink (coat), mittens, nightgown, nylons, pajamas, pants, parka, robe, scarf, shawl, shirt, shoes, skirt, slippers, socks, sweater, swimsuit, tie, trousers, veil, vest
container (14)	ashtray, bag, barrel, basket, bin (waste), bottle, box, bucket, freezer, jar, mug, pot, sack, urn
fish (11)	catfish, cod, goldfish, guppy, mackerel, minnow, perch, salmon, sardine, trout, tuna
fruit (29)	apple, avocado, banana, blueberry, cantaloupe, cherry, coconut, cranberry, grape, grapefruit, honeydew, lemon, lime, mandarin, nectarine, olive, orange, peach, pear, pineapple, plum, prune, pumpkin, raisin, raspberry, rhubarb, strawberry, tangerine, tomato
furniture (15)	bed, bench, bookcase, bureau, cabinet, chair, couch, desk, dresser, lamp, rocker, shelves, sofa, stool (furniture), table
herbivore (18)	buffalo, caribou, cow, deer, elk, fawn, giraffe, gopher, grasshopper, hare, moose, otter, ox, porcupine, sheep, skunk, squirrel, turtle
insect (13)	ant, beetle, butterfly, caterpillar, cockroach, flea, grasshopper, hornet, housefly, moth, spider, wasp, worm
mammal (57)	bat (animal), bear, beaver, bison, buffalo, bull, calf, camel, caribou, cat, cheetah, chimp, chipmunk, cougar, cow, coyote, deer, dog, dolphin, donkey, elephant, elk, fawn, fox, giraffe, goat, gopher, gorilla, groundhog, hamster, hare, horse, hyena, lamb, leopard, lion, mink, moose, mouse, otter, ox, panther, pig, platypus, pony, porcupine, rabbit, raccoon, rat, seal, sheep, skunk, squirrel, tiger, walrus, whale, zebra
musical instrument (18)	accordion, bagpipe, banjo, cello, clarinet, drum, flute, flute, guitar, harmonica, harp, harpsichord, keyboard (musical), piano, saxophone, trombone, trumpet, tuba, violin
pet (22)	bat (animal), budgie, canary, cat, chickadee, dog, finch, goldfish, guppy, hamster, hare, iguana, mink, mouse, parakeet, pigeon, pony, python, rabbit, rat, salamander, turtle
predator (17)	alligator, cat, cheetah, cougar, coyote, eagle, falcon, fox, hawk, hyena, leopard, lion, mink, owl, panther, tiger, vulture

Appendix A
(Continued)

Superordinate	Basic-Level Exemplars
tool (34)	axe, bolts, broom, brush, chain, chisel, clamp, comb, corkscrew, crowbar, drill, fork, hammer, hatchet, hoe, ladle, level, microscope, paintbrush, pencil, pliers, rake, sandpaper, scissors, screwdriver, screws, shovel, sledgehammer, spade, spear, stick, tomahawk, wheelbarrow, wrench
utensil (22)	bowl, broom, colander, corkscrew, cup, dish, fork, grater, hatchet, knife, ladle, mixer, mug, paintbrush, pan, pen, pencil, pot, spatula, spoon, strainer, tongs
vegetable (31)	asparagus, avocado, beans, beets, broccoli, cabbage, carrot, cauliflower, celery, corn, cucumber, eggplant, garlic, lettuce, mushroom, olive, onions, parsley, peas, pepper, pickle, potato, pumpkin, radish, rhubarb, rice, spinach, tomato, turnip, yam, zucchini
vehicle (27)	airplane, ambulance, bike, boat, bus, canoe, car, cart, dunebuggy, helicopter, jeep, motorcycle, sailboat, scooter, ship, skateboard, submarine, tank (army), tractor, trailer, train, tricycle, trolley, truck, van, wagon, yacht
weapon (39)	armor, axe, bat (baseball), baton, bayonet, bazooka, belt, bomb, bow (weapon), cannon, catapult, crossbow, crowbar, dagger, grenade, gun, hammer, harpoon, hatchet, hoe, knife, machete, missile, pistol, revolver, rifle, rock, rocket, shield, shotgun, shovel, sledgehammer, slingshot, spear, stick, stone, sword, tomahawk, whip

Appendix B: Superordinate concept-feature pairs used in feature verification (Experiment 2 and Simulation 2)

Superordinate	Feature	Feature Type (Wu & Barsalou, 2008)
appliance	is hot	Internal surface property
	used for cooking	Function
	is electrical	Internal surface property
bird	builds nests	Entity behavior
	lays eggs	Entity behavior
	has feathers	External component
carnivore	lives in forests	Location
	is fast	Systemic property
	has a tail	External component
clothing	made of cotton	Made of
	worn for warmth	Function
	different colors	External surface property
container	used for carrying things	Function
	has a lid	External component
	made of plastic	Made of

Appendix B
(*Continued*)

Superordinate	Feature	Feature Type (Wu & Barsalou, 2008)
fish	lives in lakes	Location
	is smelly	External surface property
	swims	Entity behavior
fruit	has a pit	Internal component
	tastes good	Internal surface property
	tastes sweet	Internal surface property
furniture	used for relaxing	Function
	has drawers	Internal component
	made of wood	Made of
herbivore	has horns	External component
	eats grass	Entity behavior
	has 4 legs	External component
insect	lives in the ground	Location
	crawls	Entity behavior
	is small	External surface property
mammal	has legs	External component
	eats	Entity behavior
	is brown	External surface property
pet	lives in cages	Location
	has a long tail	External component
	flies	Entity behavior
predator	hunts	Entity behavior
	has claws	External component
	has eyes	External component
tool	used for carpentry	Function
	has a handle	External component
	made of metal	Made of
utensil	is long	External surface property
	is round	External surface property
	found in kitchens	Location
vegetable	is crunchy	Internal surface property
	is nutritious	Systemic property
	is edible	Function
vehicle	is loud	External surface property
	has an engine	Internal component
	used for transportation	Function
weapon	is sharp	External surface property
	used for war	Function
	used for killing	Function

Appendix C: Prime–target pairs for superordinate priming (Experiment 3 and Simulation 3)

	Superordinate Prime		
Typicality	Related	Unrelated	Exemplar Target
High	appliance	mammal	toaster
	bird	vehicle	budgie
	carnivore	utensil	crocodile
	clothing	bird	jacket
	container	appliance	bucket
	fruit	container	pear
	furniture	insect	bookcase
	insect	weapon	beetle
	mammal	tool	donkey
	tool	fruit	hammer
	utensil	carnivore	spoon
	vegetable	clothing	spinach
	vehicle	vegetable	car
	weapon	furniture	knife
Low	appliance	mammal	radio
	bird	vehicle	penguin
	carnivore	utensil	dog
	clothing	bird	gloves
	container	appliance	basket
	fruit	container	coconut
	furniture	insect	bench
	insect	weapon	flea
	mammal	tool	chipmunk
	tool	fruit	brush
	utensil	carnivore	mixer
	vegetable	clothing	mushroom
	vehicle	vegetable	airplane
	weapon	furniture	rocket