# On the Nature and Scope of Featural Representations of Word Meaning

Ken McRae
University of Western Ontario

Virginia R. de Sa
University of Rochester

Mark S. Seidenberg
University of Southern California

Behavioral experiments and a connectionist model were used to explore the use of featural representations in the computation of word meaning. The research focused on the role of correlations among features, and differences between speeded and untimed tasks with respect to the use of featural information. The results indicate that featural representations are used in the initial computation of word meaning (as in an attractor network), patterns of feature correlations differ between artifacts and living things, and the degree to which features are intercorrelated plays an important role in the organization of semantic memory. The studies also suggest that it may be possible to predict semantic priming effects from independently motivated featural theories of semantic relatedness. Implications for related behavioral phenomena such as the semantic impairments associated with Alzheimer's disease (AD) are discussed.

Many theories have assumed that word meaning is represented, at least in part, in terms of featural primitives (see, e.g., Collins & Quillian, 1969; Minsky, 1975; Norman & Rumelhart, 1975; Shallice, 1988; and Smith & Medin, 1981, for overviews). Several properties of such representations have been explored in detail and have proven to have explanatory value. For example, a number of studies have shown that different features are activated depending on the context in which the word occurs (Barsalou, 1982), suggesting that word meanings are not like fixed dictionary entries. The purpose of the present research was to examine the role of featural representations in the processing of word meaning. Three general issues were addressed: the relevance of featural representations to different types of semantic tasks; the nature of featural representations, focusing on the way in

which feature correlations might be learned and what their subsequent role in word recognition might be; and the organization of semantic memory, with particular emphasis on defining semantic relatedness and specifying the source of automatic semantic priming.

## THE SCOPE OF FEATURAL REPRESENTATIONS

The validity of a featural approach to word meaning has been widely questioned in recent years. Much contemporary research on concepts has focused on higher level knowledge, such as people's naive theories of biology (for discussion, see Jones & Smith, 1993, and associated commentaries). Knowledge-based theories (Medin, 1989; Murphy & Medin, 1985) attempt to account for phenomena such as the development of conceptual structures (Keil, 1989) and people's capacity to reason about category membership (Rips, 1989). In this approach, higher level knowledge is assumed to be central to concepts and is positioned as an alternative to feature-based accounts. However, the two views seem to address different phenomena: Theories seem irrelevant to recognizing and reacting appropriately to everyday objects, whereas features cannot account for people's performance in tasks such as Rips' that involve explicit reasoning based on conceptual representations. The present research examined the hypothesis that whereas featural representations are central to the initial computation of word meaning, they are less relevant to the kinds of tasks that typically lend support to knowledge-based theories.

Jones and Smith (1993) have noted that the tasks used in studies of conceptual representation may yield data about numerous aspects of human knowledge and processing because they vary greatly in terms of the information required to perform them. The tasks that have been used to assess people's concepts can be viewed as spanning a continuum that extends from high-level, untimed reasoning tasks to

speeded tasks that rely primarily on what is sometimes termed automatic processing (Posner & Snyder, 1975). Tasks situated toward the high-level end are those that demand considerable reasoning, for example, assessing the relative probabilities that two patients have a certain disease (Medin, Altom, Edelson, & Freko, 1982); assessing whether an object that was a horse but was given surgery to look like a zebra, and now acts like a zebra but still gives birth to horse babies is really a horse or a zebra (Keil, 1989); or assessing whether a circular object that is two inches in diameter is a pizza or a quarter, in the absence of any other information (Rips, 1989). Performing these types of judgments requires integrating a number of information sources, such as the perceptual features of objects, knowledge of one's language, general knowledge of biology, and knowledge of the pragmatics of answering questions in psychology experiments. Toward the low-level end are speeded object naming, feature verification, same–different judgments, and semantic decisions (e.g., "is it a concrete object?"). Performance on these tasks may depend largely on featural representations, with minimal influence from higher level reasoning processes. Tasks that fall somewhere between these extremes include untimed similarity, concept typicality, and feature typicality judgments. Although these tasks involve explicit comparisons among multiple aspects of stimuli, they require less extensive reasoning than those typically employed by researchers studying knowledge-based theories.[1]

Studies that have examined the informational bases of tasks such as analogical reasoning, similarity judgments, and speeded same–different judgments provide support for this view (Goldstone, 1992; Markman & Gentner, 1993; Medin, Goldstone, & Gentner, 1993). For example, Markman and Gentner used the structure-mapping theory of analogy (Gentner, 1983) to investigate the bases of similarity judgments. Their experiments suggest that making explicit, untimed judgments of the similarity of two scenes involves finding a globally consistent mapping between them and then assessing feature overlap with respect to the global alignment. In a related study, Goldstone used a deadline procedure to investigate the factors that influence fast and slow same–different judgments. When subjects were required to respond in one second or less, same–different judgments were dominated by local, direct, matching of individual features. In contrast, when they were allowed 2.68 s to respond, they used the time to align the stimuli before assessing feature overlap, as in the Markman and Gentner studies.

We extended this line of research by using two pairs of yoked tasks.[2] All four tasks used words as stimuli and made use of feature norms collected in Experiment 1. Experiment 2 contrasted a short stimulus onset asynchrony (SOA) semantic priming task (2A) with an untimed similarity rating task (2B). Experiment 3 contrasted speeded feature verification (3A) with untimed feature typicality rating (3B). Each pair of tasks was assumed to differ in terms of degree of cognitive penetrability (Pylyshyn, 1984), and it was hypothesized that featural measures would better predict

subjects' performance on the speeded, less cognitively penetrable tasks.

## THE NATURE OF FEATURAL REPRESENTATIONS OF WORD MEANING

In order to make detailed predictions regarding the influence of featural representations on the computation of word meaning, a specific theoretical framework was required. Connectionist models present a relevant framework in that all models that use distributed representations incorporate featural representations of some sort. Because of this, connectionist models have previously been used to investigate the acquisition, representation, and use of featural semantics. For example, research by Hinton (1981) used featural representations to explore word meaning and the structure of categories. Hinton and Shallice (1991) and Plaut and Shallice (1993) used a similar approach to explore deep dyslexia, a behavioral impairment that is marked by a number of symptoms, the most salient of which is a tendency to make semantic errors in reading aloud (e.g., reading *symphony* as *orchestra*). Such models are trained to compute a featural representation of a word or picture using a standard learning algorithm such as backpropagation. Exposure to a large set of stimulus patterns results in the encoding of information about the distribution of features across concepts. For example, a model might learn not only that a *tiger* ⟨has fur⟩ but also that ⟨has fur⟩ is a feature of many other animals and that animals that ⟨have fur⟩ also tend to ⟨have claws⟩ and ⟨have a tail⟩.[3] This property of connectionist networks highlights an important question that is a focus of this article: Do people learn feature correlations, and if so, what role does this knowledge play in computing word meaning?

Insight into the role of correlated features in word recognition can be gained by considering the nature of the lexical mappings among spelling, sound, and meaning. The mapping from English spelling to sound can be characterized as a system of soft regularities in that words with similar spelling tend to have similar pronunciation (Seidenberg & McClelland, 1989). A major reason for the success of the Seidenberg and McClelland model of word naming is that the backpropagation learning algorithm is particularly well-suited to encoding such quasi-regular mappings. In contrast to the spelling-sound mapping, a classic computational

---

[1] This type of classification of tasks has been fruitfully used in a number of other areas of cognitive psychology. For example, Roediger and McDermott (1993) classified tests of implicit memory along a perceptual–conceptual dimension. Their scheme has been viewed as extremely useful to researchers interested in integrating results from different yet related methodologies (e.g., Moscovitch, Goshen-Gottstein, & Vriezen, 1994).

[2] In actual fact, Experiments 2 and 3 predated these studies; see McRae (1991).

[3] When referring to entities in the world, lowercase normal font is used. Concept names are printed in italics, as are examples of experimental stimuli. Names of features are printed in angled brackets as in ⟨has fur⟩.

problem in lexical semantics is that the mapping from word form (spelling or sound) to meaning is unsystematic in many languages, particularly for monomorphemic words. Thus, similar word forms (e.g., *hat, cat, mat,* and *that*) must be mapped onto dissimilar semantic patterns. This problem can also be expressed in terms of the lack of reliability of subword mappings: Letters do not map onto specific components of meaning.

Some researchers such as Forster (1994) have taken these facts to indicate that connectionist systems are ill-suited for computing word meaning. This view is mistaken for several reasons. First, it overstates the limitations on the capacity of connectionist networks to learn arbitrary mappings. Feedforward networks, which were Forster's focus, can learn arbitrary mappings if provided with sufficient numbers of hidden units. Networks that are allowed to memorize a set of patterns sacrifice the capacity to generalize, but this is irrelevant when the mapping between domains is arbitrary. Second, it overstates the arbitrariness of the relationship between form and meaning. Although the mapping between codes (e.g., phonology $\Rightarrow$ semantics) is largely arbitrary (for monomorphemic words), the structure that exists within each type of representation is also relevant. Both phonology and semantics are highly constrained domains in which only some combinations of features are valid and features are correlated with one another to different degrees. In fact, the structured nature of semantics played a central role in the Hinton and Shallice (1991) and Plaut and Shallice (1993) accounts of deep dyslexia.

The model presented below belongs to a class of connectionist architectures that use a correlational learning algorithm to encode covariations among features that then form the basis for processing; these are known as attractor networks. From this class, we chose a Hopfield network (1982, 1984). Unlike the strictly feedforward backpropagation networks that are more common in cognitive psychology, the processing of a Hopfield network is iterative and quite transparent. Thus, when an input pattern is presented, activation over a Hopfield net's representational units changes over time until a learned pattern is computed (where a learned pattern is a basic level concept in our case). That is, over a number of iterations (or processing cycles), the network settles into (or converges on) a stable state that represents one of its learned patterns. For any individual stable state, there exists a set of similar representational states that collectively constitutes its basin of attraction. A basin of attraction can be visualized in three dimensions as an irregularly shaped sink with a learned pattern located at its lowest point (i.e., the drain hole). When a Hopfield network gets into a basin of attraction and no input intervenes, it settles into the corresponding stable state.

How does this work in the case of computing a word's meaning from its spelling or sound? Because the mapping from word form to meaning is largely arbitrary, the first processing cycle produces a pattern over the word-meaning units that falls within the word's basin of attraction, but only roughly approximates the correct meaning. Then, assuming that each unit in the network represents a semantic feature and a correlational learning algorithm like the Hopfield rule

(1982, 1984) has been used, knowledge of the correlations among semantic features transforms the pattern over a number of processing cycles until it corresponds to the word's meaning. Thus, when each unit corresponds to a feature, knowledge of feature correlations is a major influence in determining the number of iterations required for the network to converge.

This understanding of attractor network dynamics motivated a major prediction tested in the experiments presented below, that correlated features influence the speed with which word meaning is computed. We also assessed the further prediction that such effects would be more apparent in speeded tasks (Experiments 2A and 3A) than in slower untimed tasks (Experiments 2B and 3B). Finally, we implemented a Hopfield network in order to assess the validity of these predictions about its behavior. Because of the complexity of interacting, nonlinear systems, it is necessary to assess an implemented model rather than rely on conjectures about its performance. This was particularly important in the present case because we were relying on the model to encode knowledge of correlated features in the weights (that is, as part of the processing mechanism), rather than adding specific representational units for this purpose (see Gluck, Bower, & Hee, 1989).

The modeling described next also improved on previous efforts in that it was not based on the type of ad hoc featural representations that have been employed in the work on word meaning. Instead, our featural representations were empirically motivated, being based on the normative data described in Experiment 1. One benefit of this approach was that it allowed us to discover that the distributional patterns of correlated features differed across artifacts and living things, a pattern that was predicted by Gelman (1988) and Keil (1989), who stressed that living things cohere around clusters of correlated features, but artifacts do not. Based on these analyses, Experiment 2 tested for a dissociation between living things and artifacts in terms of the influence of correlated features on automatic priming effects. In addition, the simulations allowed us to investigate whether the network would naturally distinguish between living things and artifacts on this basis, a fact that has been recently used to model category-specific deficits in Alzheimer's dementia (Devlin, Gonnerman, Andersen, & Seidenberg, 1996).

## CORRELATED FEATURES

Before proceeding, it is important to clarify what we mean by *correlated features* in this article because the term has been used imprecisely in the concepts literature. In the present research, two features were correlated if they tended to appear in the same basic-level concepts (see Experiment 1 for the method used to compute these correlations).[4] For example, according to the correlational analyses on the norms described below, the features ⟨has fur⟩ and ⟨has

---

[4] Some of the concepts may not be at the basic level because they included 19 birds and *bird* is often considered to be basic level. A number of analyses suggested that excluding the birds did not alter the results.

whiskers) are significantly correlated because living things like tigers, dogs, and lions that have fur also tend to have whiskers.

In contrast, in the artificial concepts literature, investigations of correlated features have involved testing whether subjects can discriminate between categories on the basis of pairs of features rather than single features (e.g., Medin et al., 1982; Waldmann, Holyoak, & Fratianne, 1995; Wattenmaker, 1991, 1993). The correlated features in those studies differ from the present work in both their form and purpose. Murphy and Wisniewski (1989) have pointed out that this literature has used correlated features to refer to what are actually conditional probabilities. Rather than measuring how features are correlated across the concepts that a person has learned, these experiments focus on a pair of features that pattern differently in two categories, thus providing a reliable cue for categorization. Consequently, features correlated at +1 in one category and −1 in another may be key to distinguishing between categories in a study such as Medin et al., but would be measured as independent in our study. It might also be noted that it is unclear whether any pair of natural categories has this +1, −1 structure (e.g., in one category, all exemplars either ⟨have wings⟩ and ⟨have a beak⟩ or have neither, but in the other category, all exemplars either ⟨have wings⟩ or ⟨have a beak⟩, but not both). Because of these key differences, the experiments and associated models that deal with configural cues were not emphasized in the present article (e.g., Gluck et al., 1989; Hintzman, 1986; Kruschke, 1992; Medin & Schaffer, 1978).

The term correlated features has most commonly been used to refer to the correlations that are hypothesized to form the basis of basic-level concepts (Rosch, 1978). For example, a person's concept of *dog* might consist of the set of features that are strongly correlated over the individual dogs that she has encountered. These correlations are learned by encoding the features that occur with objects that have been labeled *dog*. This knowledge was important in our work because subjects drew on it when listing features in the norms of Experiment 1. These norms then formed the foundation for computing feature correlations across concepts.

Only one study has specifically investigated the link between correlated features across basic-level concepts and word meaning (Malt & Smith, 1984). Using representations based on feature norms, Malt and Smith computed the correlation between each feature pair and tested for an influence of correlated features on typicality judgments. They found that augmenting Rosch and Mervis' (1975) family resemblance measure with information about correlated features improved predictions of typicality ratings slightly, but nonsignificantly. This result is consistent with our view of task effects on the influence of correlated features; the effect of correlated features found by Malt and Smith was small because typicality rating is a relatively slow judgment task that is insensitive to the time course of computing word meaning and is potentially influenced by a number of information sources relevant to knowledge-based theories.

## SEMANTIC RELATEDNESS

The semantic-priming paradigm has played a major role in explorations into how semantic memory is organized and seems particularly suited to investigating whether correlated features influence the dynamics of computing word meaning. The semantic-priming task and the notion of semantic relatedness have been used in numerous studies of language processing in normals (see Neely, 1991, for a recent review), as well as studies of Alzheimer's dementia (e.g., Nebes, Brady, & Huff, 1989), amnesia (e.g., Shimamura & Squire, 1984), and implicit memory (e.g., Schacter, 1992). Semantic relatedness is typically contrasted with associative relatedness. Robust effects of associative relatedness have been found in a large number of experiments (Neely, 1991). The widely accepted view is that associative relatedness results from temporal contiguity in speech or text (McKoon & Ratcliff, 1992), or word co-occurrence within a proposition (McNamara, 1992).

Semantic relatedness is less clear for two reasons: There are conflicting empirical results, and the theoretical basis of semantic relatedness has not yet been established. A few studies have worked toward understanding the factors underlying semantic relatedness priming. Using simultaneous presentation of prime and target, Fischler (1977) found semantic priming with associative relatedness removed (when associative relatedness was operationalized by word-association norms). The source of Fischler's priming effect is questionable, however. In a lexical decision task, if the target is related to the prime, it must be a word (nonwords have no a priori semantic relations). Neely and Keefe (1989) have found evidence suggesting that priming effects in lexical decision are influenced by subjects retrospectively checking whether the prime and target are related. Thus, it is probable that Fischler's results were actually due to large retrospective effects that stemmed from the double lexical decision task. In a more recent study, Hodgson (1991) examined the bases of semantic priming at a range of short SOAs and found consistent effects with a lexical decision task. Unfortunately, he informed subjects of the related nature of some of the stimuli, a procedure that likely encouraged strategic processing. Furthermore, Hodgson's effects were small and homogeneous across the types of relationships that he studied (category coordinates, antonyms, synonyms, subordinate–superordinate pairs, conceptual associates, and phrasal associates), prompting him to claim that the effects were due to retrospection.

Two recent studies, Shelton and Martin (1992, Experiment 4) and Moss, Ostrin, Tyler, and Marslen-Wilson (1995, Experiment 3), used single presentation to avoid strategic effects and found little or no priming. In fact, even prior to Moss et al.'s replication, Shelton and Martin claimed that "words that are very similar in meaning or sharing many features will not show automatic semantic priming if they are not also associated" (p. 1204). These failures are confusing because all existing theories of automatic priming predict the opposite result. In fact, these results prompted Shelton and Martin to conclude that automatic priming does not involve semantic representations at

all, but is due instead to associative relationships at the word-form level. However, Shelton and Martin chose their prime–target pairs based on their intuition that the concepts possessed sufficient featural overlap to produce a priming effect. (Moss et al. chose their items based on their intuition that the primes and targets were category coordinates.) This reliance on intuition is particularly problematic when conclusions are based on null results, as was the case in their studies. In fact, Lund, Burgess, and Atchley (1995) have commented on the apparent lack of relatedness between some of the pairs used by Shelton and Martin, such as *duck–cow*, *knife–hammer*, and *magazine–record*. Similarly, Moss et al. also used several pairs that did not seem to possess a high degree of featural overlap, such as *lake–mountain*, *pig–horse*, and *kite–balloon*. Further support for this notion comes from a similarity-rating task conducted by McRae and Boisvert (1997) in which the prime–target pairs of Experiment 2 of this article were rated as more similar (6.4, on a scale of 1 to 9) than those of Shelton and Martin (3.6) and Moss et al. (4.5). Thus, it appears that their null priming effects resulted from using prime-target pairs that were not sufficiently similar. In other words, their experiments were not a strong test of the prediction that semantic relatedness is due, at least in part, to featural similarity of word meaning.[5]

This study combined the feature norms of Experiment 1 with the connectionist network in order to produce a representational and processing framework that enabled us to take a step toward defining semantic relatedness and to make detailed predictions. Specifically, semantic relatedness was measured as overlap between the prime and target in terms of individual features and correlated feature pairs. Featural similarity was then used to predict priming effects on an item-by-item basis.

## SUMMARY

In summary, the reasoning processes that underlie tasks such as similarity judgments can involve many types of knowledge, only one of which is the featural representation made available when word meaning is computed. Therefore, variables that govern the representation of features in semantic memory might be expected to exert strong effects on tasks that are closely tied to word recognition, but weak effects on tasks that require considerable reasoning. This hypothesis was examined by conducting experiments in which the same stimuli were used with two types of tasks; one task required using featural information that rapidly becomes available in the course of recognizing words, whereas the other encouraged subjects to reason about the stimuli and permitted them to use nonfeatural types of knowledge in making their response. Within this framework, we tested the hypothesis that effects of correlated features should be evident only in the speeded tasks. Furthermore, the influence of featural representations on automatic semantic priming was tested in a more detailed way than in past research by using similarity in terms of individual and correlated features to account for item-by-item priming effects. Following the experiments, connectionist

simulations are presented to illustrate how correlated features might be learned and how they might have influenced the tasks of Experiments 2 and 3.

## EXPERIMENT 1

The purpose of Experiment 1 was to obtain data concerning subjects' knowledge of concepts by having them generate features that were then used to construct representations for a large set of words. This feature-listing method has been used in many previous studies (e.g., Barsalou, Olseth, & Wu, 1996; Rosch & Mervis, 1975; Smith, Osherson, Rips, & Keane, 1988). The resulting norms are assumed to provide valid information not because they yield a literal record of semantic representations but rather because such representations are systematically used by subjects when generating features. They therefore provide a window into important aspects of word meaning without necessarily being definitive (Medin, 1989). We further restricted the study to words that refer to concrete objects, specifically 19 concepts from each of the following 10 categories: *birds, mammals, fruit, vegetables, clothing, furniture, kitchen items, tools, vehicles,* and *weapons*. Subjects were given a form containing 20 concept names (e.g., *dog, desk*) and were asked to list features for them (e.g., for *dog:* ⟨barks⟩, ⟨has paws⟩). Space was provided for 10 features. Subjects were asked to take a couple of minutes per concept to list as many features as possible.

### Method

#### Subjects

Three hundred McGill University undergraduate and graduate students participated. Thirty subjects listed features for each concept. Their names were entered into a lottery for a cash prize, or, in some cases, they were paid $2.

#### Materials

Four living-thing categories (*birds, mammals, fruits,* and *vegetables*) and six artifact categories (*clothing, furniture, kitchen items, tools, vehicles,* and *weapons*) were used (see Appendix A). Battig and Montague (1969) and Rosch (1975) served as guides for choosing exemplars that spanned a range of typicality within each category. There were nine exemplar pairs in each category; these served as the similar prime–target pairs in Experiment 2. In addition, in order to facilitate the regression analyses of Experiment 2, the degree of similarity among those pairs varied somewhat (even though all were still part of the similar group). That is, based on intuition, some pairs were virtually synonymous (e.g., *sofa–couch*), whereas others were only moderately similar (*carrot–celery*). In addition, a prototypical exemplar was paired with the superordinate category name.

---

[5] The term "featural similarity" is used in this article to denote a measure of similarity computed over individual and correlated features. Wherever we intend to refer specifically to similarity in terms of one or the other type of featural representation, it is specified.

For the feature-listing task, there were 10 forms containing 20 concept names, 2 from each category. No form contained both members of a prime–target pair, and each contained 1 superordinate category name. Subjects were asked to list features of the things to which the words referred. They were asked to list different types of features, such as "physical (perceptual) properties (how it looks, sounds, smells, feels, and tastes); functional properties (what it is used for; where and when it is used); and encyclopedic facts (such as where it is from, or historical facts)." Three examples were provided. Thus, each subject listed features for 20 concepts, and 30 subjects listed features for each concept. Subjects were given as much time as needed; they took approximately 40 min on average.

## Procedure

*Distribution and collection.* Initially, 500 forms were distributed in McGill University psychology classes, and 167 were returned (33%). In the second round, 200 forms were distributed, and 80 were returned (40%). The final 53 forms were collected in the laboratory.

*Recording the features.* For each concept, each feature was recorded with its production frequency, which is the number of subjects who listed the feature for that particular concept (between 1 and 30). The initial stage in analyzing the norms was to ensure that synonymous features were recorded identically, both within and between concepts. For example, ⟨used for transportation⟩, ⟨used for transport⟩, ⟨is used for transportation⟩, ⟨people use it for transportation⟩, and ⟨transportation⟩ were coded as ⟨used for transportation⟩. It was equally important to ensure that features differing in meaning were given unique labels. Responses were interpreted conservatively, and the validity of all but the most obvious interpretations were verified with three naive colleagues. The following method was adopted for interpreting and organizing the feature set. Unless it was deemed important, quantifiers such as *generally, usually,* and *can be* were dropped. It was assumed that information carried by these quantifiers was inherent in production frequency data; that is, the number of subjects who included a feature in their lists should vary according to how often the instances have the feature. If a subject listed an adjective–noun feature (e.g., ⟨has 4 wheels⟩), it was divided (⟨has wheels⟩ and ⟨has 4 wheels⟩) under the assumption that the subject had provided 2 bits of information.[6] Disjunctive features (e.g., ⟨is green or red⟩) were also divided (into ⟨is green⟩ and ⟨is red⟩). A number of key words and phrases were used to organize and code the features. These are displayed in italics, with an example completion in normal font: ⟨*a* deer⟩ (synonym), ⟨*bought/sold* in hardware stores⟩, ⟨*causes* gas⟩, ⟨*eaten* for dessert⟩, ⟨*eg*—jeans⟩, ⟨*found* in kitchens⟩, ⟨*grows* underground⟩, ⟨*has* paws⟩, ⟨*is* brown⟩, ⟨*isa* tool⟩, ⟨*lives* in the forest⟩, ⟨*made* of metal⟩, ⟨*part of a* table setting⟩, ⟨*requires* a driver⟩, ⟨*runs on* gasoline⟩, ⟨*used* for carpentry⟩, and ⟨*worn* by women⟩.

## Results and Discussion

We first provide some general descriptive statistics regarding the data set. Two contrasting representations of word meaning are then described, one based on individual features and the other on correlated feature pairs. These representations were the basis for the experiments and modeling described below. The final set of analyses illustrate

that correlated features tend to be more dense for living things than for artifacts.

Before proceeding, note that the superordinate concepts (e.g., *mammal*) tended to be treated differently than the exemplar concepts (e.g., *horse*). Subjects commonly remarked that they found it "hard" or "strange" to list features for the superordinates. In fact, subjects failed to provide coherent feature descriptions for eight of them; *kitchen* and *bird* were exceptions presumably because *kitchen* is a place rather than a superordinate category name (*kitchen items* was actually the category), and *bird* is a basic-level concept for many adults (Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976). Subjects' inability to produce systematic features for superordinates can be illustrated by the percentage of responses consisting of category exemplars (e.g., listing *truck* for *vehicle*). This type of response occurred frequently for *mammal* (17%), *fruit* (15%), *vegetable* (8%), *clothing* (14%), *furniture* (23%), *tool* (20%), *vehicle* (18%), and *weapon* (23%), but less so for *bird* (<1%) and *kitchen* (<1%). A similar result was reported in Rosch et al. Because responses to the superordinates were highly variable and unsystematic, they were excluded from all further analyses except the analyses of variance in Experiment 2A.

There were a total of 54,685 responses; each subject listed an average of 9.6 features per concept. There were 9,618 different features listed, so that each feature was listed by an average of 2.9 subjects. 2,963 features were listed for at least one concept by a minimum of five subjects. The features were categorized by type using a slightly revised version of a taxonomy developed by Barsalou et al. (1996). The results are shown in Table 1. The number of each type of feature broke down as follows: 1,640 aspects of an entity; 669 function features; 309 classifications; 294 features describing information related to a situation in which the concept takes part; and 51 features describing people's cognitions related to an exemplar of the concept. Table 1 presents the number of artifact and living thing features in each of these categories, as well as example features. Although the patterns of four of the feature types were very similar, the number of functional features was much higher for artifacts than for living things, as would be expected.

Experiments 2 and 3 were designed to examine the role of individual and correlated features on people's performance on semantic tasks. Given that all feature-based theories of semantic memory assume that individual features are represented, our analyses were designed to assess whether correlated features predicted performance above and be-

---

[6] This procedure had potential implications for computing correlations between feature pairs (see the *Correlated Features* section below). However, the number of correlated feature pairs that resulted from this procedure was minimal (14 out of the 1190 correlated feature pairs, 1%) and was approximately equal for artifacts (6, e.g., ⟨has 4 wheels⟩ and ⟨has wheels⟩) and living things (8, e.g., ⟨has a long tail⟩ and ⟨has a tail⟩). The main reason that so few correlations involved a complex feature and its simpler counterpart was that the correlational analyses involved only the 240 features that were found in 3 or more concepts, and few of the complex features were listed for greater than 1 concept.

Table 1

*Breakdown of Responses in Experiment 1 Norms*

| Category ⟨example features⟩ | Feature frequency | |
| --- | --- | --- |
| | Artifacts | Living things |
| Aspects of a concrete entity | | |
| ⟨has a handle⟩, ⟨made of wood⟩ | 852 | 758 |
| Functional information | | |
| ⟨used for carpentry⟩, | | |
| ⟨worn by women⟩ | 504 | 165 |
| Classification | | |
| ⟨isa fruit⟩, ⟨eg—jeans⟩ | 159 | 150 |
| Information related to a situation in which | | |
| it takes part | | |
| ⟨found in bedrooms⟩, | | |
| ⟨grows on trees⟩ | 156 | 138 |
| People's related cognitions | | |
| ⟨runs on gasoline⟩, ⟨is fun⟩ | 35 | 16 |

yond the influence of individual features. In order to isolate the separate influences of the two types of information, two semantic representations were constructed, one based on individual features and the other on correlated feature pairs. In the individual features representation, information about correlated features was absent because a concept was represented as a set of weighted features. In the correlated features representation, word meaning was represented in terms of correlated feature pairs, and information about individual features was not included as a distinct component.

## Individual Features Representation

Representations similar to those used by Rosch and Mervis (1975) and Tversky (1977) were constructed for the 190 exemplar concepts. A feature was included as part of a concept's representation if at least 5 of the 30 subjects had listed it. The resulting 1,242 features were represented as vectors. Each feature vector contained 190 units, with the value for $unit_{ij}$ (i = 1–190; j = 1–1,242) corresponding to the number of subjects who listed $feature_j$ for $concept_i$. Each concept was represented across the feature vectors as a 1,242-unit pattern. A sparse representation resulted; because no concept contained more than 27 features, each concept vector included a minimum of 1,215 zeros.

## Correlated Features Representation

Concepts were also represented as patterns across correlated-feature pairs. Pearson product–moment correlation (r) was computed between pairs of features. To avoid spurious correlations, only the 240 features possessed by three or more concepts were included. There were 1,190 feature pairs correlated at $p < .01$. Figure 1 is a histogram showing the frequency of correlated feature pairs as a function of percentage of shared variance. Percentage of shared variance ranged from 6.5% to 99.7%. The distribution is positively skewed with 272 of the 1,190 correlated feature

pairs sharing between 6.5% and 10% of their variance and another 466 pairs sharing between 10% and 20%.

Concepts were represented as vectors across the resulting 1,190 feature-pair units. A concept's value on a unit was determined by a three-part function. If the concept contained neither feature from the correlated pair, the unit was set to 0. If it possessed both features, the unit was set to the sum of the production frequencies. If it contained only one of the pair, then it violated the correlation, and the unit's value was set to the negated production frequency of the feature it possessed. Thus, if a concept possessed only one feature of a pair, the more often subjects had listed that feature for that concept, the stronger the violation. For example, ⟨flies⟩ and ⟨has feathers⟩ shared 43% of their variance. According to the norms, a *carrot* neither ⟨flies⟩ nor ⟨has feathers⟩ (a value of 0); an *eagle* both ⟨flies⟩ and ⟨has feathers⟩ (13 + 16 = +29), and, although 22 subjects listed ⟨has feathers⟩ for *ostrich,* none claimed that it ⟨flies⟩ (−22).

One important property of the feature vectors was that they tended to be sparse. Because of this sparsity, the presence of one feature did not reliably predict the absence of another. This can be understood by considering the Pearson correlation coefficient, which is defined as $\Sigma z_x z_y / (n - 1)$. Because the feature vectors contained predominantly zeros, mean production frequency was less than 0.4 for all features, and less than 0.1 for 96% of them. Thus, a production frequency of 0 fell close to the mean, so that an absent feature contributed little to the correlational measure and the simultaneous presence of two features dominated it. As is outlined in Appendix B, the attractor network's learning rule was similarly sensitive to pattern sparsity, and it seems reasonable that humans do not pay attention to absent
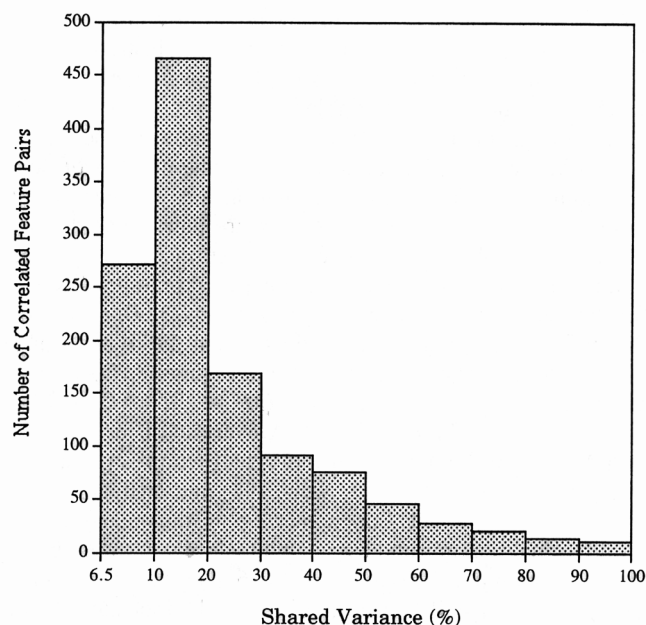


*Figure 1.* Frequency distribution of the 1,190 correlated feature pairs in terms of percentage of shared variance.

features either (i.e., because there are so many possible features in the world, the fact that a cat does not have a handle has little effect on learning).

It is also worth noting that computing feature correlations across all 190 concepts contrasts with Malt and Smith (1984), who computed correlations within superordinate category. We chose not to follow their strategy because breaking the concepts into their superordinate categories leads to a number of ambiguities, such as whether *tomato* should be included in *fruit* or *vegetable* or whether *dolphin* and *whale* should be included in *mammal*. The situation is even more problematic for artifacts because there seems to be no independent set of principles by which they can be reliably subcategorized. Thus, the appropriate categories are not at all clear, and a great deal of overlap results when possible schemes are considered (e.g., a *knife* can be a *tool*, a *utensil*, a *kitchen item*, or a *weapon*). The uncertainty is illustrated by the fact that 214 superordinate category features were listed by at least 5 subjects for the 190 concepts, greater than one per concept. Because the psychologically relevant set of superordinate categories is unclear, we felt that the best strategy was to compute correlations across all concepts.

## Artifacts and Living Things

According to Gelman (1988) and Keil (1989), living things and artifacts differ in the extent to which they entail correlated features. Living things tend to cohere around clusters of correlated features, whereas artifacts tend to cohere around the intended function of the creator (but see Malt & Johnson, 1992). The norms supported this notion in that 11% of living-thing feature pairs were significantly correlated (at $p < .01$), but only 6% of artifact pairs. In addition, living-thing concepts were more densely represented across correlated feature pairs. To illustrate this, a two-way analysis of variance (ANOVA) was conducted in which the independent variables were type of concept (76 living things versus 114 artifacts) and type of representation (individual versus correlated features). The number of positive units in a concept's representation (i.e., the number of individual features or correlated feature pairs) was the dependent variable. A significant interaction effect showed that the pattern of individual and correlated features differed for the two types of concepts, $F(1, 188) = 73.00, p < .001$. Simple main-effect analyses indicated that the living things possessed far more correlated feature pairs ($M = 33.4$, $SE = 2.1$) than did the artifacts ($M = 13.9, SE = 1.2$), $F(1, 367) = 145.88, p < .001$. However, no difference was found between the number of individual features possessed by living things ($M = 16.5, SE = 0.4$) and artifacts ($M = 15.0, SE = 0.3$), $F < 1$. For each category, the mean number of individual features and correlated feature pairs were birds (15, 33), mammals (19, 38), fruits (16, 41), vegetables (15, 21), clothing (16, 19), furniture (15, 8), kitchen items (15, 12), tools (14, 7), vehicles (15, 14), and weapons (15, 23). These analyses beg the question of whether the influence of correlated features in a word-recognition experiment might

differ between artifacts and living things, a question that is investigated in Experiment 2.

In summary, individual and correlated feature representations were constructed from the norms in order to isolate their influence on subjects' performance in Experiments 2 and 3 and to guide the connectionist modeling. The correlations were computed across concepts such as *dog, tiger, chair,* and *couch.*

## EXPERIMENT 2

The same stimuli were used in an automatic semantic priming and a semantic-similarity judgment task. These enabled us to contrast a speeded task not requiring an explicit comparison (short SOA priming) with an untimed explicit comparison task ("On a scale of 1 to 7, how similar are these 2 concepts?"). In each case, featural similarity was used to predict the dependent variable in regression analyses. Four hypotheses were tested. First, recent work has suggested that subjects use featural representations to perform speeded same–different judgments, whereas they use several sources of information when making untimed similarity judgments (Goldstone, 1992; Markman & Gentner, 1993; Medin et al., 1993). Thus, it was hypothesized that the featural representations would better predict performance on the priming task. Second, if correlated features are integral to computing word meaning as an attractor network would suggest, then their influence should be apparent in the speeded task that is sensitive to the temporal dynamics of the computation. Third, if featural similarity of word meaning is an important aspect of semantic relatedness, then it should predict the magnitude of priming effects. Finally, because the distribution of correlated features differs for artifacts and living things, living thing priming effects might be more strongly influenced by correlated features.

## Experiment 2A

Posner and Snyder (1975) proposed that when the SOA is short, semantic priming is automatic, whereas with more time, subjects use the prime to generate an expectancy set for the impending target. This hypothesis has since been elaborated, but the core ideas have been retained (e.g., Becker, 1980; Neely & Keefe, 1989). Further studies have established that priming effects are automatic and limited to lexical internal processing if the SOA is about 250 ms or less (Neely, 1977; Den Heyer, Briand, & Dannenbring, 1983; De Groot, 1984). Therefore, in Experiment 2A, a prime, such as *lamp*, was presented 250 ms prior to the onset of a target, such as *chandelier*. Subjects were instructed to read the prime and to make a decision to the target. Rather than the lexical decision task that is standard in priming experiments, semantic decision tasks such as "is it animate?" were used for two reasons. First, whereas semantic decisions are clearly based on semantic knowledge, the lexical decision task affords other bases for making responses and therefore does not necessarily demand use of semantics (Balota & Chumbley, 1984; Seidenberg &

McClelland, 1989). Second, because semantic decisions were used, fillers were constructed that nullified retrospective priming (i.e., when a response is influenced by a subject evaluating the prime–target relationship). Any influence of retrospective checking can be eliminated in a semantic decision task such as "is it animate?" by including related but inanimate filler pairs (e.g., *house–cottage*); if the number of these fillers equals the number of related pairs for which the response to the target is "yes," relatedness no longer cues the response. The particular choice of semantic decision is also important. Experiment 2A featured broad semantic decision tasks in order to discourage subjects from generating specific exemplars of the category. For example, it is unlikely that subjects would spend time generating exemplars of *living things* or *things made by humans* because these categories have so many members. Narrow semantic decision or categorization tasks (e.g., "is it a bird") are inappropriate because they may cue specific exemplars, thereby altering the basis of obtained effects (Jared & Seidenberg, 1992).

## Method

### Subjects

Forty-eight McGill University undergraduates were paid $3 each. Twenty-four subjects were randomly assigned to each list. None had participated in Experiment 1.

### Materials

The semantically similar prime–target items consisted of 90 pairs of concepts varying in degree of featural similarity. They are presented in Appendix A, organized in pairs of columns so that the primes are directly left of their corresponding targets. Also included were the 10 superordinate concepts paired with 10 typical exemplars. Experiment 2 was conducted simultaneously with the collection of the feature norms, and it was not known at that time that the norms would fail to render coherent concepts for the superordinate categories. Therefore, the superordinate–exemplar pairs were included in Experiments 2A and 2B, and the ANOVA of 2A, but were excluded from the regression analyses.[7]

There were four semantic decision tasks: "is it animate?", "is it an object?", "is it made by humans?", and "does it grow?". The similar prime–target pairs for the "is it animate?" task were birds and mammals, and the dissimilar primes were taken from among the tools and clothing (e.g., *eagle–hawk* vs. *sandals–hawk*). The similar prime–target pairs for the "does it grow?" task were fruits and vegetables, and the dissimilar primes were taken from among the furniture, kitchen items, vehicles, and weapons (e.g., *lettuce–cabbage* vs. *stereo–cabbage*). The similar prime–target pairs for the "is it an object?" task were furniture, kitchen items, vehicles, and weapons, and the dissimilar primes were taken from among the birds and mammals (e.g., *pistol–rifle* vs. *cow–rifle*). The similar prime–target pairs for the "is it made by humans?" task were tools and clothing, and the dissimilar primes were also taken from among the birds and mammals (e.g., *shoes–boots* vs. *dog–boots*).

The test and filler trials were designed so that relatedness did not cue the response in any way. In each task, there were 50% "yes" and 50% "no" trials. The probability of a "yes" response following a "yes" or "no" prime was also 50% (although subjects did not respond to the prime). Furthermore, because semantic decisions were used, related "no" trials could be included with the same frequency as related "yes" trials (e.g., *square–triangle* in the "is it animate?" task). The proportion of related trials was 0.25, and half of the unrelated trials required a "no" response. In addition, there were 8 lead-in practice trials balanced in the same way. There was a separate "is it an adjective?" practice session that contained 24 trials, also balanced in the same way.

Two lists were constructed for each task so that subjects saw each target only once. Each list contained half of the targets with similar primes and the opposing half with dissimilar primes. Subjects were informed that a recognition test would be administered following the practice session and each task; this was included to encourage them to attend to the primes. Five primes were used as recognition test stimuli, and 5 words not appearing in any task were used as foils.

### Procedure

In Experiments 2A, 2B, and 3A, presentation of stimuli and recording of responses were accomplished using Micro Experimental Laboratory (MEL) software running on an IBM XT-286 microcomputer (International Business Machines, Armonk, NY). Each subject performed the four tasks in a blocked fashion. Subjects were told that all words were meant to be nouns. Each priming trial proceeded as follows: an intertrial interval of 1,500 ms; an asterisk in the center of the screen for 250 ms; a 250-ms pause; a prime for 200 ms; a mask that consisted of &&&&&&&&& for 50 ms; and the target until a response was made. The experimenter asked subjects to read the first word (the prime) silently and to respond as quickly and accurately as possible to the second word (the target). Examples of positive and negative exemplars were provided: animate (i.e., something that is alive; e.g., a cockroach is but a keyboard is not); object (i.e., something that is tangible or concrete; e.g., a pair of scissors is an object, but the sky is not); made by humans (i.e., something that is manufactured by people; e.g., a razor but not a butterfly); and grows (i.e., something that grows on its own; e.g., a spider grows, but a door does not). A session began with the practice trials. For the recognition tests that followed the practice and each task, subjects were asked to indicate whether or not each of 10 items had appeared as a prime. It took approximately 40 min for subjects to complete the experiment.

### Results

Latency and accuracy of responses were recorded. All response latencies greater than three standard deviations from the mean of the correct test trials in each task were replaced by the cutoff value. The percentage of response latencies affected by this procedure were 1.4% for "is it

---

[7] *Emu–ostrich* and *starling–crow* were also removed from the regressions because it was apparent that few people knew the meanings of *emu* or *starling*. In Experiment 2b, 11 of 40 subjects reported being completely unfamiliar with *emu*, and 8 were unfamiliar with *starling*. Furthermore, in Experiment 1, 13 of 30 subjects either failed to list any features for *emu* or listed incorrect ones such *has fur;* 10 of 30 subjects listed inappropriate features for *starling*. Finally, in separate norms in which subjects rated the familiarity of the concepts on a scale of 1 to 7, the mean rating for *emu* was 2.45 (190th of 190) and for *starling* was 2.95 (184th of 190).

animate?"; 1.5% for "is it an object?"; 2.4% for "is it made by humans?"; and 0.6% for "does it grow?". Overall, subjects made errors on 2.2% of similar trials and 3.7% of dissimilar trials. By task, the errors were "is it animate?", similar = 1.9%, dissimilar = 5.8%; "is it an object?", similar = 1.9%, dissimilar = 2.9%; "is it made by humans?", similar = 3.1%, dissimilar = 3.0%; and "does it grow?", similar = 2.3%, dissimilar = 4.0%). The errors were not further analyzed due to their low frequency.

## Overall Results

Although the regressions were the primary focus of attention, ANOVAs were also conducted in order to illustrate that significant overall priming effects were obtained. There were two factors: prime type, with 2 levels: similar and dissimilar; and task, with 4 levels: "is it animate?", "is it an object?", "is it made by humans?", and "does it grow?". In the analyses by subjects, both task and prime were within subjects. In the item analyses, prime type was within items, and task was between.

Overall, decision latency was faster for targets preceded by a similar prime ($M = 760$ ms, $SE = 10$ ms) than by a dissimilar prime ($M = 807$ ms, $SE = 11$ ms), $F_1(1, 46) = 36.32$, $p < .0001$, and $F_2(1, 96) = 31.98$, $p < .0001$. Decision latency also differed by task, $F_1(3, 138) = 47.72$, $p < .0001$, and $F_2(3, 96) = 34.80$, $p < .0001$. Mean latency by task was animate, 723 ms ($SE = 14$ ms); grow, 752 ms ($SE = 10$ ms); object, 758 ms ($SE = 9$ ms); and made by humans, 926 ms ($SE = 17$ ms). No interaction between task and prime type was apparent, $F < 1$ by subjects and items.

Planned comparisons showed that decisions were faster for targets preceded by similar primes in three of the four tasks: "is it animate?", 692 ms ($SE = 17$ ms) versus 755 ms ($SE = 21$ ms), $F_1(1, 184) = 13.59$, $p < .0004$, and $F_2(1, 192) = 6.02$, $p < .02$; "is it an object?", 733 ms ($SE = 11$ ms) versus 783 ms ($SE = 12$ ms), $F_1(1, 184) = 10.35$, $p < .002$, and $F_2(1, 192) = 7.59$, $p < .007$; and "is it made by humans?", 899 ms ($SE = 23$ ms) versus 952 ms ($SE = 24$ ms), $F_1(1, 184) = 7.19$, $p < .009$, and $F_2(1, 192) = 4.26$, $p < .05$. However, the priming effect in the "does it grow?" task was not significant, 741 ms ($SE = 13$ ms) versus 764 ms ($SE = 15$ ms), $F_1(1, 184) = 3.52$, $p < .07$, and $F_2 < 1$.

## Regression Analyses

Featural similarity was computed as the cosine of the angle between two concept vectors. In the individual features representation, each of the 190 concepts was a vector of production frequencies across 1,242 features. Cosine between pairs of concept vectors increased linearly with number of shared features and decreased linearly with number of distinct features. A cosine of 1 corresponded to identical concepts, and 0 corresponded to concepts that shared no features. For example, using the individual features representation, sofa and couch had a cosine of .879, pineapple and coconut had a cosine of .415, screwdriver and drill had a cosine of .214, and chair and whale were

orthogonal (cosine = 0). In the correlated features representation, a concept was a vector across 1,190 correlated feature pairs. Thus, cosine increased with number of shared correlated pairs and decreased with number of distinct pairs. It also decreased when one of the concepts violated a correlation, but the other did not. Thus, a negative cosine was possible because violated correlations were represented by negative integers. According to the correlated features representation, sofa and couch had a cosine of .963, pineapple and coconut had a cosine of .219, screwdriver and drill had a cosine of .061, and chair and whale were again orthogonal (cosine = 0).

Prime–target similarity in terms of individual features was used to predict priming effects on an itemwise basis by predicting mean response latency for related prime–target pairs (lamp–chandelier) after mean response latency for unrelated pairs (goose–chandelier) had been forced into the regression equation (i.e., had been partialed out). Similarity in terms of correlated features was then used to predict the residual variance. Thus, by first entering similarity in terms of individual features, the burden of proof was placed on showing an influence of correlated features. Because the analyses on the norms revealed that correlated features were more prominent for living things, regressions were conducted separately for living things and artifacts.[8]

Similarity in terms of individual features predicted priming effects for artifacts, $r^2 = .15$, $F(1, 51) = 8.96$, $p < .005$, but not for living things, $r^2 = .04$, $F(1, 31) = 1.30$, $p > .3$. Conversely, similarity in terms of correlated feature pairs predicted priming effects for living things, $r^2 = .25$, $F(1, 30) = 10.19$, $p < .004$, but not for artifacts, $r^2 = .003$, $F < 1$.[9]

---

[8] Performing regressions in this manner essentially treats subjects as a fixed effect (Clark, 1973). The regressions do, nonetheless, represent a strong test because the independent variables were based on norms that were conducted using a separate sample of subjects, and, did not, in any way, mention similarity among concepts. Furthermore, no norming subject saw both members of a prime-target pair. The same arguments hold for the regression analyses of Experiment 3 because the subjects of the norming experiment were distinct from those of the feature verification and feature typicality rating experiments.

[9] Our measures of semantic similarity, though significant predictors of priming effects, accounted for only about 25% of the variance. Predictive accuracy may have been reduced if it was the case that subjects were making implicit semantic decisions to the primes. Taking the "is it animate?" task as an example, the primes for the similar test trials were animate (e.g., caribou–moose), but the primes for the dissimilar control trials were inanimate (e.g., drill–moose). Decision stage latency may thus have been reduced for similar targets, resulting in an uniform increment of the priming effect across the prime-target pairs. If this was the case, predictive accuracy in the regression analyses would have been reduced (or unaffected) because the correlations depended on featural similarity for the similar, but not for the dissimilar, pairs. That is, predictive accuracy depended on the variation in featural similarity among only the similar prime-target pairs. The critical point here is that all similar primes possessed the decision feature (i.e., was animate, was an object, was made by humans, or was grown). Therefore, the fact that the similar, but not the dissimilar,

In order to better understand the factors involved in predicting living-thing priming effects, the four main parameters of the correlated features representation were manipulated. These parameters were including only significantly correlated feature pairs rather than all pairs; representing feature-pair violations; restricting the analysis to features that occurred in at least three concepts; and representing each significantly correlated feature pair by a single vector, regardless of the strength of its relationship.

Table 2 summarizes the regression analyses. Each regression used a representation in which all parameters except the highlighted one were identical to the correlated features representation of Experiment 1. First, the priming effects were not predicted when the representation included all feature pairs, rather than being restricted to correlated ones, $F < 1$.[10] This result is consistent in spirit with the configural cue model of Gluck, Bower, and Hee (1989). In their model, there would be a large weight between a feature-pair node and a category node only if the co-occurrence pattern of the feature pair is informative for category membership. That is, although the model includes a configural cue node for each feature pair, only informative pairs influence processing to any degree. This result is also consistent with models such as Medin and Schaffer (1978) and Hintzman (1986), in which the major factor that determines the degree to which a feature pair influences the similarity metric is the extent to which they co-occur. The similarity metric, in turn, directly affects categorization.

Second, when information about violated correlations was removed from the representation, prediction was no longer significant, $p > .09$. Therefore, important information is carried by the feature pairs that violate the correlational structure, as well as those that obey it. Third, when the correlated features representation was constructed from those that occurred in two or more concepts, predictions were unaffected, $p < .006$. However, if correlated feature pairs were constructed from all of the features, predictions were nonsignificant ($p > .2$) because of the many spurious correlations involving features that occurred in only one concept. The final analysis involved weighting pairs by how strongly they were correlated, which might conceivably increase predictive accuracy. To construct such a representation, significantly correlated pairs were represented by 1–10 vectors, contingent upon amount of shared variance. Strength of correlation was coded discretely by using it to determine the number of units assigned to a feature pair using the following rule: for one, $.064 \leq r^2 < .164$; for two, $.164 \leq r^2 < .264$; for three, $.264 \leq r^2 < .364$; and so on. Predictions using this scheme were similar to the unweighted version of the correlated features representation, $p < .009$. Finally, all of these representations predicted less than 1% of the residual variance of artifact-priming effects.

Table 2

*Manipulating the Parameters of the Correlated Features Representation When Predicting Human Priming Effects for Living Things*

| Manipulated parameter | $r^2_{partial}$ | $F(1, 30)$ | $p$ |
|---|---|---|---|
| Original representation | .25 | 10.19 | <.004 |
| Including all pairs, correlated or not | .01 | <1 | |
| Removing violation information | .09 | 2.93 | >.09 |
| Including features that were part of 2 or more concepts | .23 | 9.04 | <.006 |
| Including features that were part of 1 or more concepts | .04 | 1.38 | >.2 |
| Weighting with strength of correlation | .21 | 8.00 | <.009 |

*Note.* Each manipulation involved changing only the named parameter; all other parameters matched the original representation.

## Nonlinearities

One surprising result was that similarity in terms of individual features failed to predict priming effects for living things. This suggests that if correlated features are particularly dense in some part of semantic space, featural similarity is not adequately captured by a linear combination of individual features. There are three factors that may have caused the individual features similarity measure to fail to predict the living-thing priming effects. First, because the measure was linear with respect to number of shared features, it did not behave as a positively accelerating function (see Shepard, 1958, 1987, for a discussion of the nonlinear generalization–similarity gradient). A number of models have incorporated mechanisms that transform exemplar similarity into a positively accelerating function (e.g., Hintzman, 1986; Medin & Schaffer, 1978; Neuman, 1974). In Neuman's model, for example, storing an item in memory involved keeping a count of shared features and feature pairs. Therefore, if concepts shared three features, they had 6 units in common; if they shared four features, they had 10 units in common; five features, 15 units, and so on. Thus, each additional shared feature had a greater impact on similarity, producing a positively accelerating gradient, rather than a linear one.

Second, similarity in terms of individual features increases with shared features and decreases with distinct ones. Likewise, similarity in terms of correlated features increases with shared pairs and decreases with distinct ones. A point of divergence is where a shared individual feature decreases similarity in terms of correlated features. If one

---

primes were animate (or an object, or made by humans, or grown) may have adversely affected the regression results, but probably did not influence them at all.

[10] One other difference between the individual and correlated feature representations was that the correlated feature representation was based only on the 240 features possessed by three or more concepts. To ensure that the results were not due to this, the regressions were repeated with similarity in terms of those 240 individual features replacing the correlated features measure. Prediction of residual priming effects with this individual features representation was nonsignificant for artifacts, $r^2 = .03$, $F(1, 50) = 1.53$, $p > .2$, and living things, $r^2 = .004$, $F < 1$.

concept obeys a correlation but the other violates it, similarity in terms of correlated features is lower than if the second concept possessed neither feature. Because the regression analyses revealed that information about correlational violations was important for predicting priming effects, we analyzed the number of cases where one concept obeyed a correlation but the other violated it. The mean number of these cases per concept was greater for living things ($M = 21.3$, $SE = 1.7$) than for artifacts ($M = 8.4$, $SE = 1.1$), $t(86) = 6.82$, $p < .0001$, suggesting that living things may have been more affected by this second departure from linearity.

A third possibility is that similarity in terms of individual features failed to predict living-thing priming effects simply because its variance was too low. This seems unlikely for two reasons. First, in Experiment 2B below, similarity in terms of individual features predicted living-thing similarity judgments. Second, the standard deviations of the two measures patterned identically for artifacts and living things. That is, for both types of concepts, the standard deviation of the similarity measure was larger for correlated features (artifacts: $M = .382$, $SD = .313$; living things: $M = .400$, $SD = .231$) than for individual features (artifacts: $M = .522$, $SD = .207$; living things: $M = .571$, $SD = .185$).

In summary, because both correlated feature pairs and violations of them were more dense for living things, featural similarity was not adequately captured by a linear measure over individual features (i.e., independent cues). In contrast, because the features of artifacts tend to be sparsely intercorrelated, individual features were sufficient. Further discussion is deferred until after Experiment 2B.

## Experiment 2B

A separate set of subjects were shown the similar prime–target pairs and were asked to rate the similarity of the things to which the words referred on a scale of 1–7. There were two main hypotheses. First, when the computation of word meaning is viewed in terms of an attractor network, correlated features influence the speed of the computation (i.e., rate of convergence to a stable state). Because rating concept similarity is not a speeded task, correlated features should not predict similarity ratings, even for living things. Second, the degree to which individual features should predict similarity ratings for either artifacts or living things is uncertain because other sources of information influence these judgments.

### Method

#### Subjects

Forty McGill University undergraduates were paid $3 each. None had participated in Experiments 1 or 2A.

#### Materials

List 1 contained the concept pairs in one order (e.g., *rug–mat*) and List 2 contained them in the reverse order (e.g., *mat–rug*).

Both orderings were used because Whitten, Suter, and Frank (1979) have reported order effects in a similarity-rating task. Each list contained all 90 exemplar pairs, the 10 superordinate–exemplar pairs, and 20 fillers. The fillers ranged from very dissimilar (e.g., *ant–tripod*) to slightly similar (e.g., *beach–sandbox*) in order to pull subjects into the lower regions of the scale. Only a few dissimilar fillers were included because we wanted the ratings to reflect the subtle differences in the pairs of interest. Additionally, 10 lead-in practice pairs that covered a range of perceived similarity (according to the intuitions of the first author) were used to orient subjects to the task.

### Procedure

Word pairs were presented on a computer screen, and subjects were asked to rate the similarity of the things to which the words referred on a scale of "1 = not at all similar" to "7 = exactly the same thing" by pressing an appropriate number on the keyboard. Examples were given. Subjects were told that they could take their time. They were also asked to tell the experimenter if they were unfamiliar with the meaning of any word. The 10 lead-in pairs of concepts were presented in random order. Following the lead-in pairs, the 90 test and 20 filler pairs were presented randomly. Each pair of stimuli was presented on the same line, centered on the screen with five spaces separating them. There was a 1,500-ms intertrial interval. It took about 15 min to complete the experiment.

### Results

Overall, subjects reported being unfamiliar with one or both of the words on 1.5% of the trials, and these were discarded. Ratings were averaged across the two orders. No ANOVAs were conducted because there was no dissimilar condition in this experiment. For the regression analyses, featural similarity in terms of both individual and correlated features was again computed as the cosine of the angle between two concept vectors. Predicting similarity ratings differed from predicting priming effects in that no dissimilar pairs were available to act as control items. Thus, individual features were used to predict the similarity ratings, and correlated feature pairs were used to predict the residual variance. Again, regressions were conducted separately for living things and artifacts.

Similarity in terms of individual features predicted the ratings for both artifacts, $r^2 = .33$, $F(1, 52) = 26.06$, $p < .001$, and living things, $r^2 = .12$, $F(1, 32) = 4.22$, $p < .05$. In contrast, similarity in terms of correlated feature pairs predicted the ratings for neither artifacts, $r^2 = .01$, $F < 1$, nor living things, $r^2 = .03$, $F < 1$.

In Experiment 2A, individual features predicted priming effects for artifacts, and the opposite held for living things. Insofar as similarity ratings reflect individual rather than correlated features, they should predict priming effects for artifacts but not for living things. To test this, the regression analyses of Experiment 2A were repeated, except that rated similarity replaced similarity over individual features. Similarity ratings predicted priming effects for artifacts, $r^2 = .10$, $F(1, 51) = 5.50$, $p < .03$, but not for living things, $r^2 = .05$, $F(1, 31) = 1.60$, $p > .1$. In contrast, similarity in terms of correlated feature pairs predicted residual priming

effects for living things, $r^2 = .19$, $F(1, 30) = 7.17$, $p < .02$, but not for artifacts, $r^2 = .04$, $F(1, 50) = 2.14$, $p > .1$.

## Discussion

The results of Experiment 2 are important for four main reasons. First, they suggest that the mechanism that computes word meaning exploits statistical regularities among semantic features, as do attractor networks. Second, they suggest that similarity is determined by a combination of featural overlap and other knowledge, where the additional knowledge plays a larger role in slower tasks that require considerable reasoning. Third, they provide evidence that featural similarity is a primary source of short SOA priming effects and thus a primary organizing principle of semantic memory. Finally, they demonstrate that correlated features are more prominent in the on-line processing of living things than of artifacts.

## Correlated Features and Computing Word Meaning

An influence of correlated features was found in short SOA priming, but not in similarity rating. In a connectionist framework, short SOA priming can be conceptualized as reflecting interconcept distance in semantic space (for similar proposals, see Kawamoto, 1993; Masson, 1995; Plaut, 1995; Sharkey, 1989). Computing the meaning of a word can be viewed as driving semantic memory from the state that it was in prior to reading or hearing that word to the state corresponding to its meaning. In a short SOA priming task, the meaning of the prime determines the prior state for computing the meaning of the target. Because prime–target similarity in terms of individual features is a major determinant of the distance from the start to the end state, it influences the degree of facilitation. Furthermore, because feature correlations affect the manner and speed with which activation accrues, the particular pairs of features that the prime and target share also affect the speed with which the target concept is computed. In fact, because feature correlations were dense in the living things, their influence overwhelmed that of individual features. These points are discussed in more detail when the model is described following Experiment 3.

## Semantic Relatedness

Experiment 2A is the first to demonstrate that priming effects can be predicted on an item-by-item basis from an empirically derived measure of featural similarity. However, one question that needs to be addressed is whether the effects might have been due to associative relationships. There are two major arguments against this hypothesis. First, Experiment 2A did not simply demonstrate an overall priming effect using semantically related prime–target pairs. Rather, the magnitude of item-by-item priming effects was predicted by featural similarity, thus providing strong evidence that it was the basis of task performance. Second, McRae and Boisvert (1997) have recently used similarity

ratings to choose a set of prime–target pairs that were significantly more similar than those used by Moss et al. (1995) and Shelton and Martin (1992). Word-association norms showed that no prime ⇔ target associations existed. With these items, they demonstrated automatic semantic-priming effects in four conditions, short SOA and single-presentation schemes paired with semantic and lexical decision tasks. Thus, Experiment 2A, particularly in conjunction with McRae and Boisvert, demonstrates that automatic priming taps word meaning (rather than solely tapping associations at the word-form level) and that semantic relatedness can be defined, at least in part, by featural similarity.

## Aspects of Similarity

In contrast to the priming study, individual features predicted similarity ratings, but no influence of correlated features was detected. This result implies that any influence of correlated features on computing word meaning was masked by the leisurely paced and relatively complex similarity-rating task. In Experiment 2B, similarity in terms of individual features predicted the ratings quite well. In fact, for the artifacts, it predicted 33% of the variance in the similarity ratings, versus 15% of the variance in the priming data. Further analyses showed that knowledge of taxonomic relationships among animals played an important role in the ratings, consistent with the knowledge-based theories of Gelman and Wellman (1991) and Medin (1989). Consider some of the mammal stimuli, such as cow–bull and deer–fawn. According to the norms, the features of cow include ⟨is female⟩, ⟨is docile⟩, and ⟨produces milk⟩, but those of bull include ⟨is male⟩, ⟨is aggressive⟩, and ⟨has horns⟩, so that the resulting cosine was .249, which was the 79th highest of the 88 concept pairs. In contrast, people know that cows and bulls are different genders of the same animal; consequently, their mean similarity rating of 5.4 was 25th highest of the 88 pairs. A post hoc analysis of the four living-thing categories reinforced the notion that similarity judgments between mammals were based on criteria outside the realm of the featural representations. The proportion of variance of the similarity ratings accounted for by individual features was: birds (17%), fruit (19%), vegetables (15%), and mammals (1%).

## Living Things Versus Artifacts

Why do the features of living things tend to be more densely intercorrelated than those of artifacts? Consider the constraints on the structure of the two types of objects. The structure of living things is determined by genetic–evolutionary principles; plants and animals have evolved over time into their present form. Correlated sets of features have evolved in parallel and become instantiated in a number of plants and animals. For example, it might be expected that sets of features have become instantiated in a number of animals that live in a specific environment. A mammal that hopes to survive in a cold environment like the Canadian

north most likely has fur, padded feet, and warm blood. In contrast to living things, artifacts are created by humans and, as such, are subject to constraints imposed by society. Typically, artifacts are designed to fulfill a specific function and must be esthetically and economically attractive to potential consumers. Other factors that may affect the structure of an artifact include the availability and cost of materials, present-day fads, and various whims of the designer and manufacturer. Because of these rather arbitrary constraints, the structural features of artifacts tend to be much more variable than those of living things, with the result that features tend to be less densely intercorrelated across artifact concepts (see Gelman, 1988, and Keil, 1989, for related arguments). Surface color serves as a particularly salient example of the arbitrariness of many artifact features. Humphrey, Goodale, Jacobson, and Servos (1994) have shown that color is a salient cue when people name living things, but not when they name artifacts. This presumably occurs because color is less variable in living things (e.g., most bananas are yellow, most moose are brown) and often carries important information (e.g., bananas turn from green to yellow to brown, and it is important to know what this means). In contrast, the color of an artifact is often arbitrary (a chair can be painted any color), although social conventions do produce tendencies for some objects to be certain colors (e.g., toasters tend to be silver or white), and color is sometimes tied to function (stop signs are painted red to be perceptually salient). In summary, the features of living things may be more densely intercorrelated because the constraints tend to be more arbitrary for artifacts than for living things.

It is also possible that people explicitly perceive living things to have a greater number of correlated features, so that any real-world difference is exaggerated in mental representations of both an explicit and implicit nature. It may be that mental representations of artifacts contain fewer correlated features because people treat them as items that are designed to perform a single function. Consequently, although the features of certain artifacts may be intercorrelated, people may be less likely to store this information because a single feature is receiving the bulk of their attention. In contrast, people tend to treat living things as complex beings that express many potential behaviors and have many potential functions. Consequently, people may attend to many features of each living thing, and this attentional difference might increase the likelihood of implicitly and explicitly encoding feature co-occurrences. Empirical support for this notion has been obtained by Billman (1989) and Billman and Knutson (1996) who have demonstrated that feature correlations are easier to learn when they are part of a system of correlations. In summary, due to attentional factors, representational differences between artifacts and living things may be exaggerated relative to objective real-world differences.

## EXPERIMENT 3

A second pair of yoked tasks, speeded feature verification (3A) and untimed feature typicality rating (3B), were used

to investigate two main hypotheses. First, featural representations should better capture performance in speeded verification than in feature typicality rating because the rating task is relatively slow and involves complex decision processes (e.g., "Should I rate this feature as a 5 or a 7?"). Second, if correlated features play a role in the computation of semantic representations, their influence should be strong in speeded verification, but not in feature typicality rating.

Thirty-seven features that were correlated with a number of others were designated as *target features*, and each was paired with two concepts. The concepts were chosen so that production frequency of the target feature was equated between groups; this measure represented the strength of association between the concept and feature according to the individual features representation. The groups differed in that the features of one of the concepts were strongly intercorrelated with the target feature, and the features of the other were weakly intercorrelated with it. For example, the target feature ⟨hunted by people⟩ was correlated with 16 others. As shown in Figure 2, *deer* contained 11 of those features, but *duck* contained only 4. These items enabled two sets of analyses. First, if the temporal dynamics of computing semantic representations depend on correlated features, ⟨hunted by people⟩ should be verified more quickly for, but not rated as more typical of, *deer* than *duck*. However, more importantly than the *t* tests that this design permitted, regression analyses were conducted on the 74 items (37 target features by two concepts each). As in Experiment 2, the key test was whether the correlated features measure (intercorrelational strength, as described below) predicted variance above and beyond that of the individual features measure (production frequency from Experiment 1). In addition, seven other independent variables were used to predict feature verification latencies in Experiment 3A and typicality ratings in 3B. These regression analyses provided a more complete picture of the
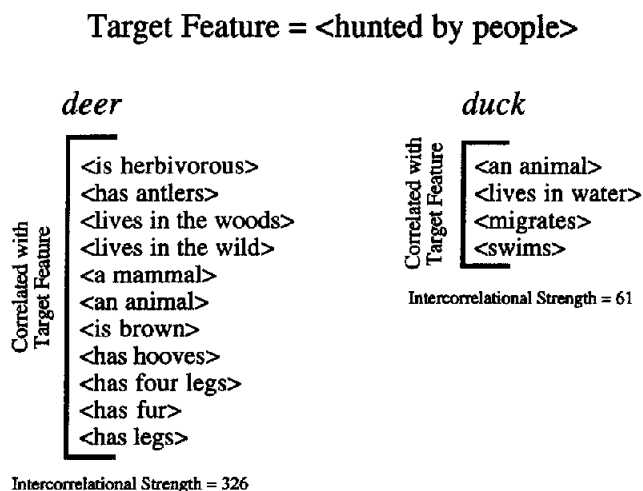
## Target Feature = ⟨hunted by people⟩



*Figure 2.* Example of an item from Experiment 3. The target feature, ⟨hunted by people⟩, was more highly correlated with features of *deer* (strong group) than of *duck* (weak group).

factors that underlie the relationship between a concept and its features and the relevance of these factors in different task environments. Experiment 3 extended Experiment 2 in that the tasks involved decisions to specific features, thus providing a more direct test of the influence of correlated features. In addition, it tested the efficacy of a continuous measure of the strength of correlation between feature pairs.

## Experiment 3A

In the feature verification task, a concept name, such as *deer*, was presented for 400 ms, followed by a target feature name, such as ⟨hunted by people⟩. Subjects were asked to indicate, as quickly and accurately as possible, whether or not the target feature was reasonably true of the concept. For example, the correct response would be "yes" to *deer*–⟨hunted by people⟩, but "no" to *horse*–⟨has scales⟩. A "reasonably true" rather than an "always true" criterion was necessary because, for example, in an "always true" decision task, *apple*–⟨is red⟩ and *knife*–⟨is sharp⟩ would require "no" responses.

### Method

#### Subjects

Twenty McGill university undergraduates were paid $2 each. None had participated in Experiments 1 or 2.

#### Materials

There were 74 concepts, each with an associated target feature (see Appendix C). The main independent variable, intercorrelational strength, was continuous and was defined as the summed shared variance between the target feature and the other features within the concept. Only feature pairs from the correlated features representation were included. Each target feature appeared with two concepts. The intercorrelations were stronger for the "strong" group ($M = 175.4$, $SE = 12.0$) than for the "weak" group ($M = 19.7$, $SE = 3.6$), $t(36) = 14.42$, $p < .0001$ (one-tailed). An example target feature, ⟨hunted by people⟩, is shown in Figure 2 with its associated concepts *deer* and *duck*. According to the norms, ⟨hunted by people⟩ is more strongly intercorrelated with features of *deer* (intercorrelational strength = 326) than of *duck* (intercorrelational strength = 61).

A number of aspects of the stimuli thought to be relevant to concept–feature relationships were equated. Production frequency, which roughly reflects a feature's accessibility when a concept is computed from its name, was equated strong ($M = 10.5$, $SE = 0.7$), weak ($M = 10.4$, $SE = 0.8$), $t(36) = 0.35$, $p > .4$ (two-tailed). In addition, verification latency or typicality rating may be influenced by a feature's salience as defined in relation to other features of the concept. This influence might be independent of the feature's absolute production frequency. Therefore, features were ranked within each concept on the basis of production frequency, and target feature rank was equated across groups: strong ($M = 8.3$, $SE = .7$), weak ($M = 8.1$, $SE = .7$), $t(36) = 0.23$, $p > .8$ (two-tailed). It was also thought that familiar concepts might be computed more quickly or have higher asymptotic levels of activation, thus speeding verification latencies as in the word-frequency effect (McRae, Jared, & Seidenberg, 1990). Furthermore, subjects might be more comfortable rating features of familiar concepts, resulting in systematically higher ratings. Familiarity was operationalized by having a separate group of 20 subjects judge the familiarity of the "thing that the word refers to" on a 7-point scale. This variable was equated between groups, strong ($M = 5.4$, $SE = 0.2$), weak ($M = 5.3$, $SE = 0.2$), $t(72) = 0.01$, $p > .9$ (two-tailed). Finally, in regression analyses designed to predict sentence-verification latencies (e.g., "robin has wings"), Ashcraft (1978) found that the total number of features produced for a concept in a norming task was a primary predictor, presumably because it reflects the ease with which a concept's features can be accessed. Therefore, this factor was also roughly equated (strong: $M = 298$, $SE = 4$; weak: $M = 287$, $SE = 5$), $t(72) = 1.70$, $p > .09$ (two-tailed). Although the difference might be considered marginally significant, it is less than 1/3 of a feature per concept per subject. In addition, this variable was not a significant predictor in Experiment 3A or 3B.

In summary, 74 stimuli were constructed that varied on a number of dimensions considered potentially relevant to concept–feature relationships. Regression analyses were used to investigate the factors that influenced performance on feature verification and feature typicality rating. In addition, 37 pairs of matched stimuli allowed for a categorical test of the influence of correlated features. The strong and weak groups differed in terms of intercorrelational strength and were equated in terms of production frequency, ranked production frequency, concept familiarity, and number of features produced for that concept.

Note that the items were not segregated on the basis of artifact versus living thing in this experiment. Instead, we took advantage of the fact that there do exist artifact features that are highly intercorrelated. In order to select a sizable set of items, target features were chosen so that they were correlated with a number of others, regardless of whether they were associated with artifact or living-thing concepts. Thus, the artifact–living thing distinction was neither manipulated nor relevant in the present experiment.

Two lists were constructed so that no target feature was presented twice to a subject. Although 13 concepts appeared with 2 target features, repetitions were assigned to different lists so that subjects saw each of these concepts only once. However, because *parakeet* appeared with 3 features, it appeared twice in List 1. Also, because *carpet* appeared with 4 features, it appeared twice in each list. List 1 contained 18 items from the strongly intercorrelated group and 19 items from the weak group. List 2 contained 19 strong items and 18 weak ones. List 1 contained 16 living things and 21 artifacts. List 2 contained 15 living things and 22 artifacts. Thirty-seven filler items were included for which the feature was not reasonably true of the concept (e.g., *tangerine*–⟨is silver⟩). Type of feature (e.g., part, function, characteristic action) was approximately matched between test features and fillers to prevent subjects from using it as a cue to their response. An additional 10 fillers (5 positive, 5 negative) served as lead-in practice items. A unique set of 20 items (10 positive, 10 negative, type of feature matched) were used in a separate practice session.

#### Procedure

Subjects were instructed to press the *yes* key (always beneath their dominant hand) as quickly and accurately as possible if the feature was reasonably true of the concept or press the *no* key if it was not. The reasonably true response criterion was explained to them, and an example was provided. Following the practice session, 10 lead-in items were presented in random order, followed by the test items and fillers, randomly ordered. Each trial proceeded

as follows: a 1,500-ms intertrial interval; an asterisk in the center of the screen for 500 ms; a blank screen for 100 ms; a concept name for 400 ms; and the target feature until the subject responded. The experiment lasted approximately 5 min.

## Results

### Overall Results

Latency and accuracy of responses were recorded. All latencies greater than 3 standard deviations from the mean of the correct test trials were replaced by the cutoff value (1.5% of the trials). Two-tailed paired samples $t$ tests showed that subjects were faster to judge that a feature was part of a concept if it was strongly correlated with other features possessed by that concept ($M = 820$ ms, $SE = 38$ ms) than if it was weakly correlated ($M = 912$ ms, $SE = 46$ ms), $t_1(19) = 4.22$, $p < .001$, $t_2(36) = 4.23$, $p < .001$. Subjects committed errors on 8.6% of the test trials, and 4.2% of the negative filler trials. Using the square root of the number of errors as the dependent variable (Myers, 1979), it was found that more errors were committed in the weak condition ($M = 13\%$) than in the strong condition ($M = 4.3\%$), $t_1(19) = 4.34$, $p < .001$, $t_2(36) = 3.24$, $p < .003$.

### Regression Analyses

Regression analyses were conducted in order to obtain a detailed picture of the relative influence of a number of predictor variables on feature verification latency (as well as feature typicality rating in Experiment 3B). There were nine independent variables, including the five that were described earlier and were considered most likely to influence verification latency or typicality rating: intercorrelational strength, production frequency, feature rank, concept familiarity, and total responses per concept. The sixth variable was the number of letters in the feature name, a factor that influences reading time but is not relevant to the rating task. This variable was automatically equated in the $t$ tests because the same features appeared in both groups. The sev-

enth independent variable was the number of features per concept from the individual features representation, which might influence a feature's activation through "gang" effects (McClelland & Rumelhart, 1981). The eighth variable, feature superordinate typicality, was measured as the number of concepts in a superordinate category that contained that feature. For example, because five *mammal* concepts possessed ⟨hunted by people⟩, it received a feature–superordinate typicality score of five for *deer*. A spillover effect of this variable might be possible in that the feature's typicality at one level of the conceptual hierarchy might trickle down, thereby influencing ratings or verification latencies (Clapper & Bower, 1991). Finally, although Ashcraft (1978) found that concept typicality was not a primary predictor of verification latency, it was included by virtue of its ubiquitous use in concept experiments.

A stepwise regression revealed that the best equation to predict verification latency included intercorrelational strength, concept familiarity, and feature rank, in that order. This equation predicted 40% of the variance. Table 3 provides a comprehensive picture of the correlations among all the variables (it also includes feature typicality rating from Experiment 3B). Table 4 presents the results of predicting feature verification latency with each independent variable. Intercorrelational strength was the best predictor (18%). Concept familiarity, feature rank, and production frequency were also reliable. Critically, intercorrelational strength predicted verification latency over and above production frequency ($r^2 = .19$, $p < .001$) and feature rank ($r^2 = .17$, $p < .001$), the two individual features measures. Table 5 shows partial correlations that represent benchmarking an equation containing each of the listed variables against one containing those variables plus intercorrelational strength. Thus, each partial correlation represents the predictive ability of intercorrelational strength over and above the listed variables. Critically, intercorrelational strength predicted a significant proportion of residual variance at each step in these analyses. Thus, not only was a measure of correlated features the best predictor of verification latency, it accounted

Table 3
*Correlations Among Feature Verification Latency, Feature Typicality Rating, and the Nine Independent Variables of Experiment 3*

| Variables | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. Verification latency | — | −.32* | −.42* | −.29* | .30* | −.33* | −.21 | −.21 | • | • | • |
| 2. Feature typicality rating | −.32* | — | .22 | .30* | • | • | .28* | • | • | • | −.20 |
| 3. Intercorrelational strength | −.42* | .22 | — | • | • | • | .38* | .32* | .21 | .26* | • |
| 4. Production frequency | −.29* | .30* | • | — | −.79* | • | • | • | • | • | −.21 |
| 5. Feature rank | .30* | • | • | −.79* | — | • | • | • | .36* | .32* | • |
| 6. Concept familiarity | −.33* | • | • | • | • | — | • | • | .20 | • | • |
| 7. Concept typicality | −.21 | .28* | .38* | • | • | • | — | • | .39* | .32* | • |
| 8. Feature–superordinate typicality | −.21 | • | .32* | • | • | • | • | — | −.22 | −.20 | • |
| 9. Features per concept | • | • | .21 | • | .36* | .20 | .39* | −.22 | — | .64* | • |
| 10. Total responses per concept | • | • | .26* | • | .32* | • | .32* | −.20 | .64* | — | • |
| 11. Letters in feature name | • | −.20 | • | −.21 | • | • | • | • | • | • | — |

*Note.* • indicates $|r| < .2$.
*$p < .05$.

Table 4

*Predicting Feature Verification Latency With Each Factor Separately*

| Independent variable | $r$ | $r^2$ | $F(1, 72)$ | $p$ |
|---|---|---|---|---|
| Intercorrelational strength | −.42 | .18 | 15.82 | <.0003 |
| Concept familiarity | −.33 | .11 | 8.80 | <.005 |
| Feature rank | .30 | .09 | 6.86 | <.02 |
| Production frequency | −.29 | .08 | 6.41 | <.02 |
| Feature–superordinate typicality | −.21 | .05 | 3.43 | >.06 |
| Concept typicality | −.21 | .04 | 3.32 | >.07 |
| Letters in feature name | .12 | .02 | 1.10 | >.2 |
| Total responses/concept | .07 | .00 | <1 | |
| Features/concept | −.04 | .00 | <1 | |

for a large proportion of unique variation over and above the eight other variables.

## Experiment 3B

The feature typicality rating task was conducted in a paper-and-pencil format. Subjects rated "how typical each feature is" of the corresponding concept on a scale of 1–9. As in Experiment 2B, it was hypothesized that correlated features (intercorrelational strength) should not predict performance in the untimed rating task. In addition, given the results of Experiment 2B in which individual features predicted similarity ratings, although feature typicality ratings allow for other sources of information to be used, the individual features measures (production frequency and rank) should predict them.

### Method

#### Subjects

Forty-three University of Western Ontario undergraduates participated for course credit. None had participated in Experiments 1, 2, or 3A.

### Materials

The 74 concept–target feature pairs that were used in Experiment 3A served as stimuli. Subjects rated each target feature as part of a set that included filler features. The number of filler features per concept depended on the number of target features paired with that concept. If the concept was paired with one target feature, at least four fillers were included. If it was paired with 2 target features, at least five fillers were used; greater than 2 target features, at least six fillers were used. All filler features were ones that had been listed by at least 2 subjects in Experiment 1. Because there were 316 features, the concepts were split roughly into two lists that were given to different subjects. A typical line in the rating form appeared as follows.

yacht     used on water ____

Four versions of each list were created to reduce order effects.

### Procedure

Subjects were tested in groups. Their instructions were, "Each person has a 5 page booklet that contains a number of category names. By category, I mean the set of things in the world that have that label (e.g., the set of airplanes in the world). Each category name is paired with short descriptions of a few features (e.g., *airplane has wings*). Each feature is more or less typical of the things in that category. Your task is to rate just how typical each feature is. For example, has wings is extremely typical of *airplanes* because basically all *airplanes have wings*, whereas *has a propeller* is less typical because only some *airplanes have a propeller*. Please rate the typicality of each feature on a scale of 1 to 9, where 1 = not typical, 5 = reasonably typical, and 9 = extremely typical." An example was provided. Subjects were told, "The following ratings represent my opinion; certainly, your opinion may differ. There is no 'right' or 'wrong' answer." It took about 20 min to complete the task.

### Results

#### Overall Results

As in Experiment 3A, subjects rated a feature as more typical of a concept if it was strongly correlated with other

Table 5

*Predicting Feature Verification Latency: The Unique Contribution of Intercorrelational Strength Over and Above the Listed Variables*

| Base equation | $r_{\text{partial}}$ | $r^2_{\text{partial}}$ | $F(1, 74 - k - 1)$ | $p <$ |
|---|---|---|---|---|
| fam | −.46 | .21 | 19.24 | .001 |
| fam, rank | −.48 | .23 | 20.75 | .001 |
| fam, rank, prod | −.47 | .22 | 19.93 | .001 |
| fam, rank, prod, fst | −.44 | .20 | 16.51 | .001 |
| fam, rank, prod, fst, ct | −.41 | .16 | 13.14 | .001 |
| fam, rank, prod, fst, ct, lets | −.42 | .18 | 14.16 | .001 |
| fam, rank, prod, fst, ct, lets, tr/c | −.45 | .20 | 16.17 | .001 |
| fam, rank, prod, fst, ct, lets, tr/c, f/c | −.44 | .20 | 15.52 | .001 |

*Note.* fam = concept familiarity; rank = feature rank; prod = production frequency; fst = feature–superordinate typicality; ct = concept typicality; lets = letters in feature name; tr/c = total responses per concept; f/c = features per concept; $k$ = the number of parameters in the equation, including intercorrelational strength.

features possessed by that concept ($M = 7.2$, $SE = 0.15$) than if it was weakly correlated ($M = 6.6$, $SE = 0.15$), $t_1(42) = 7.20$, $p < .0001$, $t_2(36) = 3.81$, $p < .001$.

## Regression Analyses

The regression analyses used the same 9 independent variables that were used in Experiment 3A. The dependent variable was feature-typicality rating rather than verification latency. The results of the regression analyses differed from Experiment 3A. A stepwise regression showed that the best equation to predict feature verification included production frequency and concept typicality and accounted for 16% of the variance (intercorrelational strength, concept familiarity, and feature rank predicted 40% of the variance of verification latencies in Experiment 3A). The regression analyses also showed that correlated features were not a major determinant of feature typicality ratings. From Tables 3 and 6, it is evident that production frequency and concept typicality were the only variables that predicted feature-typicality ratings. Because intercorrelational strength was not a predictor, no further analyses on it were appropriate.

If intercorrelational strength was not a predictor of the ratings, why was there a significant difference in the $t$ tests? Recall that because Ashcraft (1978) had found that concept typicality was not a predictor of verification latencies (a result that was replicated in Experiment 3A), it was not equated between groups. However, the regression analyses of Experiment 3B suggested that people tended to rate features as more typical when they were paired with typical concepts. Therefore, concept typicality may have been the major cause of the significant between-groups difference because concepts in the strong group were more typical on average, $t(36) = 2.66$, $p < .02$.

Finally, if feature-typicality ratings reflect individual features and verification latencies reflect both individual and correlated features, then intercorrelational strength should predict verification latencies when the variance accounted for by feature-typicality rating is removed. Rated feature typicality predicted verification latency, $r^2 = .10$, $F(1, 72) = 8.24$, $p < .006$, and intercorrelational strength predicted a significant proportion of the residual variance, $r^2 = .15$, $F(1, 71) = 12.29$, $p < .0008$.

Table 6

*Predicting Feature-Typicality Rating With Each Factor Separately*

| Independent variable | $r$ | $r^2$ | $F(1, 72)$ | $p$ |
|---|---|---|---|---|
| Production frequency | .30 | .09 | 6.86 | <.02 |
| Concept typicality | .28 | .08 | 6.05 | <.02 |
| Intercorrelational strength | .21 | .05 | 3.52 | >.06 |
| Letters in feature name | −.20 | .04 | 2.87 | >.09 |
| Feature rank | .16 | .03 | 1.99 | >.1 |
| Features/concept | .16 | .03 | 1.91 | >.1 |
| Total responses/concept | .14 | .02 | 1.43 | >.2 |
| Feature–superordinate typicality | .05 | .00 | <1 | |
| Concept familiarity | .02 | .00 | <1 | |

## Discussion

Correlated features predicted performance on the verification task, but not on the feature typicality rating task. Previous research by Barsalou (1987, 1989) had shown that features that are processed frequently, recently, or both as part of a concept are more likely to be accessible when the concept name is read or heard. In the present study, these factors corresponded to production frequency and rank, the individual features measures. Experiment 3A further established that the degree to which a feature is correlated with a concept's other features also affects its on-line accessibility. This is additional evidence that correlated features are central to the dynamics of computing semantic representations. Furthermore, the feature-based measures predicted verification latency better than they predicted typicality ratings, a difference that was presumably due to the processing requirements of the rating task that were not reflected by the independent variables (e.g., subjects keeping judgments internally consistent by thinking back to previous ones). It was also interesting that concept familiarity predicted verification latency but concept typicality predicted feature-typicality rating. It is not surprising that the familiarity measure predicted verification latency because familiarity (frequency) effects are ubiquitous in word recognition. However, the manner in which concept typicality might predict feature typicality is less obvious. The most plausible explanation appears to be a "halo effect" (Tversky & Kahneman, 1974); given that it is possible for a number of factors to influence complex judgments, concept typicality may have influenced feature-typicality ratings by "leaking" into them.

These results also relate to the notions of feature saliency, typicality, centrality, and diagnosticity. The frequency with which a feature is listed in a norming task (measured as production frequency or rank) has been taken to indicate saliency (Smith & Medin, 1981) or diagnosticity (Smith & Osherson, 1984). Smith and Medin defined a salient feature as one that has a substantial probability of occurring in instances of the concept. Smith and Osherson labeled a feature as diagnostic of a concept if using it to describe that concept (as in ⟨red⟩ *apple*) increases the probability that the resulting conjunctive concept is a good example of the base concept (i.e., a ⟨red⟩ *apple* is a better *apple* than is a ⟨brown⟩ one). The typicality ratings also seem to reflect feature saliency or diagnosticity because they are best predicted by the individual features measure. Medin and Shoben (1988) discussed the somewhat different case in which a feature might be listed equally often for two concepts in production norms, but might be central to one and not the other. A feature is central if it accepts little change, that is, if altering it causes a drastic change in how people view the concept. For example, Medin and Shoben found that the typicality rating for ⟨square⟩ *cantaloupe* as a *cantaloupe* was higher than for ⟨square⟩ *basketball* as a *basketball*. The notion of centrality may reflect the fact that altering a feature has repercussions for all those features that are correlated with it (or related to it). Thus, feature centrality may be determined in part by intercorrelational strength. If this

is true, the results of Experiment 3 demonstrate an empirical decoupling of feature centrality versus saliency (or diagnosticity).

Finally, Billman (1989) and Billman and Knutson (1996) have demonstrated that a feature is better learned if it is part of a system of correlated features. In the model of Billman and Heit (1988), this benefit results from focused sampling, which is a control mechanism that boosts the salience of features that participate in a detected regularity. The effect of correlated features in Experiment 3A might be viewed as resulting from years of focused sampling; features that were correlated with a number of others were activated more quickly. However, the simulation of Experiment 3 presented next shows this boost of activation without focused sampling during the learning phase.

The goal of Experiments 2 and 3 was to study the representation and processing of established adult lexical concepts. This goal led to their main strength and weakness. The main strength of these experiments was that the feature norms enabled the investigation of the influence of well-learned statistical knowledge of features and the correlations among them. This type of investigation of lexical concepts would be very difficult to conduct using artificial concepts. In fact, no researcher has attempted an artificial concepts study that would involve the type of intensive training of subjects that would enable them to encode knowledge of this sort from observation and to compute it automatically from linguistic input. On the other hand, the main weakness of studying established lexical concepts is that it demands a correlational approach because it is not possible to directly manipulate the input to subjects, as it is with artificial concepts. A number of artificial concepts experiments have found that people can encode and use knowledge of feature correlations, particularly in observational learning situations (e.g., Billman & Knutson, 1996; Wattenmaker, 1991, 1993; Younger and Cohen, 1983; but also see Murphy & Wisniewski, 1989, and Wattenmaker, 1993, for failures). Thus, the essence of the present results have been replicated in a number of concept formation studies.

## A MODEL OF COMPUTING WORD MEANING FROM WORD FORM

The remainder of the article describes a connectionist model of computing word meaning from word form. Few studies in cognitive psychology have investigated aspects of cognitive representations more complex than simple lists of features because of the difficulty in intuitively predicting the interactions between complex representations and processes. Simply, our intuitions about the behavior of complex, interactive, and nonlinear systems are often wrong, making it critical to have a computational model of the process under study. We have claimed throughout that a distributed model of semantic memory would show effects of correlated features in tasks that tap the time course of computing word meaning and described it in general terms on page 101. The goal of the modeling was to demonstrate this claim explicitly for Experiments 2 and 3. The goal was

not to advance a comprehensive theory of lexical conceptual memory; indeed, that does not seem possible given the present state of knowledge in this area. Rather, the model served as a vehicle to explicitly investigate the influence of correlated features on computing word meaning.

The model is a Hopfield network (1982, 1984) and belongs to a class of connectionist architectures in which a correlational learning algorithm encodes covariations among features that then form the basis for processing. The network represented concepts as distributed patterns of activation over units that corresponded to features from the Experiment 1 norms. Thus, the Hebb (1949) learning rule encoded information about how the features were correlated in the set of concepts that the model learned. This knowledge of feature correlations was then a major influence in determining the number of iterations required for the network to converge. Given these relatively transparent computational principles, it was felt that a Hopfield network would serve as an aid to understanding how correlated features might influence the computation of lexical concepts and, hence, people's performance on speeded semantic tasks.

A number of previous models of word recognition are related to this one. For example, Hinton and Shallice (1991) used an iterative backpropagation network in which there was a standard feedforward path from orthography to semantics through a set of hidden units, plus a loop from semantics through a set of hidden units and back. Basins of attraction were formed in the latter part of the network. Although iterative backpropagation networks can store more patterns, we used a Hopfield network because it is based on a simple correlational learning rule (Hebb, 1949). Therefore, if concepts are represented as distributed patterns over units corresponding to individual features, feature correlations are stored in a transparent manner. In contrast, because there was a layer of hidden units in the semantic loop of Hinton and Shallice's network, and because backpropagation was the learning algorithm, correlational structure was encoded in an indirect manner. Our model also differed from previous ones in that the representation of word meaning was based on empirically derived conceptual representations (the norms of Experiment 1). In previous models, concepts have been represented either by hand-crafted sets of features (e.g., Hinton & Shallice, 1991) or by random patterns of activation (e.g., Farah & McClelland, 1991; Kawamoto, 1993; Masson, 1995). This difference was critical because the goal was to investigate the effects of the distribution of features across concepts.

## Network Architecture

Figure 3 shows the model's architecture. A representation of word form was constructed by using one unit per letter triple that occurred in the 84 words that were included in the model; spaces at the beginning and end of a word were treated as characters. This resulted in 379 word-form units and provided a sparse distributed representation that roughly preserved item similarity. Because of the lack of
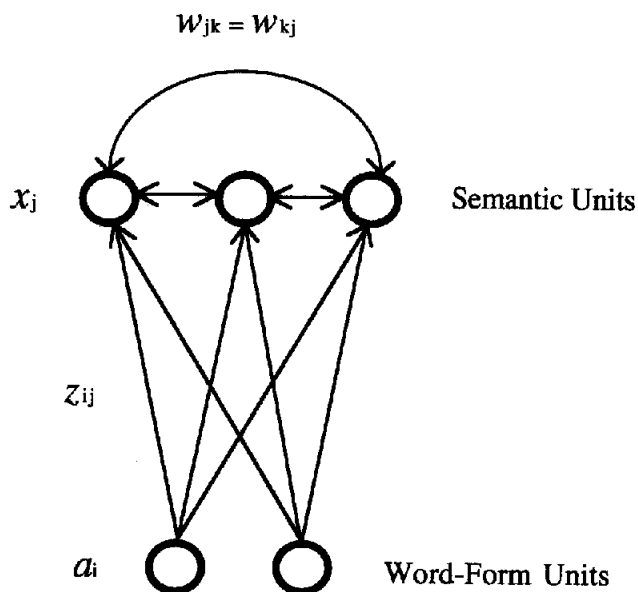
$$W_{jk} = W_{kj}$$



*Figure 3.* Basic architecture of the model (not all units are shown). The semantic units were fully bidirectionally interconnected. The orthographic units were unidirectionally connected to the semantic units and were not interconnected.

regularity in the mapping from word form to meaning, it was important that no systematic relationship existed between the input unit triples and specific features (i.e., the semantic units). An informal analysis suggested that this was the case.

The output was a distributed representation of word meaning. Each of the 646 units corresponded to a binary semantic feature (0 = feature is absent, 1 = feature is present); no production frequency information was included. Eighty-four of the 190 normed concepts were then represented as distributed patterns over the 646 feature units. The 84 concepts are shown with an asterisk (*) next to them in Appendix A. There were 10 birds, 10 mammals, 8 fruit, 10 vegetables, 8 articles of clothing, 6 pieces of furniture, 8 kitchen items, 8 tools, 8 vehicles, and 8 weapons. The 84 items included 42 prime–target pairs from Experiment 2 and 14 target features from Experiment 3. Only 84 of the 190 concepts were included in the model for computational reasons; because of the simplicity of the learning rule, Hopfield networks have limited storage capacity (Hertz, Krogh, & Palmer, 1991; Hopfield, 1982, 1984). Our sole criteria for choosing the concepts was that each category was sampled approximately equally. The four living-thing categories were slightly overrepresented to balance the six artifact categories.

The semantic units were fully bidirectionally interconnected to allow the entire range of pairwise feature correlations to be encoded. The word-form units were fully unidirectionally connected to the semantic units, but were not interconnected. Pattern processing, the learning algorithm, and initial performance analyses are described in Appendix B.

## Representations for the Model

The individual and correlated features representations described in Experiment 1 were used for Experiments 2 and 3 to choose items and predict dependent measures. However, the patterns of correlated features may have differed between the subset of 84 concepts on which the model was trained and the full set of 190 concepts. Because these patterns were of primary interest, separate representations were constructed and used to predict the behavior of the model. When correlations were computed among the 120 features possessed by a minimum of 3 concepts, 539 significantly correlated feature pairs resulted. Thus, concepts were represented across 646 individual features and 539 correlated feature pairs. These representations were used to test hypotheses in the simulations of Experiments 2 and 3. Our approach was to demonstrate that the influence of individual and correlated features on the model's behavior was analogous to their influence on human performance.

## Simulation of Experiment 2

The model was not used to simulate similarity ratings because a number of factors outside of its realm influence these judgments. Furthermore, the fact that similarity in terms of individual features contributes to similarity ratings was built into the model.

Short SOA priming can be conceptualized in terms of interconcept distance in semantic space. Computing word meaning is then viewed as moving from the representational state immediately prior to reading or hearing a word to the state that corresponds to its meaning. In short SOA priming, the prime determines the initial state for computing the target. Simulations were straightforward because of the network's temporal dynamics and sensitivity to featural similarity. Because a word's meaning is computed in 100–300 ms (Gough & Cosky, 1977; Rayner, 1978), the 250-ms SOA of Experiment 2 was assumed to be sufficient to allow a stable semantic pattern to be computed for the prime. Therefore, the prime's word form was clamped, and its meaning was computed for 10 iterations. With the prime active, the target's word form was clamped, and convergence latency was recorded. It is clear that similarity in terms of individual features partly determines amount of priming because converge latency depends on the difference between the initial and end states. In addition, an influence of correlated features may be detected because the weights directly reflect them. Returning to the sink analogy, similarity over individual features determines where the computation originates, that is, the distance from the prime to the target's basin of attraction. Feature correlations determine the time required to descend a concept's attractor basin to its stable state. Thus, both factors may influence priming. Furthermore, as with the human subjects in Experiment 2A, the influence of correlated features may be more pronounced for living things than for artifacts because the features of living things tend to be more densely intercorrelated.

To test these hypotheses, the number of iterations required for each target concept to converge was recorded when the target was preceded by a similar and a dissimilar prime (a concept that shared no features with the target). Measuring convergence is not straightforward because of the ambiguity in estimating when a concept is stable enough to support a semantic decision. Therefore, three convergence measures were used: When error dropped below 1.0, when error was within 0.1 of its value when the concept stabilized, and when error was within 0.01 of the stabilization point. The convergence measures were averaged over three runs. It was assumed that convergence latency for a concept was monotonically related to the time required for a subject to compute a word's meaning and was therefore monotonically related to how quickly and easily a subject could answer a question based on its meaning. Only 40 of the 42 priming pairs were used because the model did not possess sufficient computational resources to learn all 84 patterns (see Appendix B).

The regression analyses were identical to Experiment 2A. The dependent variable was the number of iterations for error to drop below the specified point when the target concept was preceded by a similar prime (e.g., *lamp–chandelier*). Convergence latency when preceded by a dissimilar prime (e.g., *goose–chandelier*) was the first factor forced into the regression equation. Similarity in terms of individual features was then used as a predictor. Similarity in terms of correlated features was used to predict priming effects after similarity in terms of individual features had been entered. As in Experiment 2A, similarity in terms of individual features predicted priming effects for artifacts: error less than 1, $r^2 = .27$, $F(1, 19) = 7.07$, $p < .02$; error within 0.1 of convergence, $r^2 = .59$, $F(1, 19) = 27.07$, $p < .001$; error within 0.01 of convergence, $r^2 = .58$, $F(1, 19) = 26.17$, $p < .001$; but not for living things: less than 1, $r^2 = .07$, $F(1, 15) = 1.14$, $p > .3$; within 0.1, $r^2 = .04$, $F < 1$; within 0.01, $r^2 = .01$, $F < 1$. Conversely, similarity in terms of correlated features predicted priming effects for living things: less than 1, $r^2 = .09$, $F(1, 14) = 1.39$, $p > .2$; within 0.1, $r^2 = .26$, $F(1, 14) = 4.86$, $p < .05$; within 0.01, $r^2 = .25$, $F(1, 14) = 4.69$, $p < .05$; but not for artifacts: less than 1, $r^2 = .02$, $F < 1$; within 0.1, $r^2 = .14$ in the wrong direction; within 0.01, $r^2 = .00$, $F < 1$.

A great deal of research in the past 20 years has dealt with explicating the informational bases of automatic semantic priming (Neely, 1991). Experiment 2A and its simulation demonstrate that one primary source is featural similarity. However, there are other bases as well. For example, automatic priming occurs for items that co-occur in real-world and linguistic contexts, such as *butter* and *knife*, conceptual associates, such as *dove* and *peace*, and concepts that are part of a phrase or are frequently temporally contiguous in conversation or text, such as *private* and *property*. There are two ways in which these types of priming could be captured in an attractor network. First, pattern similarity may depend somewhat on shared context in that context may be encoded as an aspect of conceptual representation (Masson, 1995). In fact, some of the features listed by subjects in Experiment 1 were contextual in nature (e.g., ⟨used for carpentry⟩, ⟨found

in bathrooms⟩). Second, Plaut (1995) has recently demonstrated that attractor networks are sensitive to sequential processing of patterns and that this sensitivity produces priming effects. Masson and Plaut have also shown that attractor networks can reproduce a variety of other priming phenomena, such as greater priming for low frequency and degraded targets, and significant priming across an interleaved unrelated item (*bread–tree–butter*). The fact that these networks can reproduce the primary empirical phenomena associated with automatic semantic priming makes them competitors of spreading activation (McNamara, 1992) and compound cue (Ratcliff & McKoon, 1988) theories. Further discussion of the relative merits of these theories can be found in Masson (1995).

## Simulation of Experiment 3

The major determinant of feature verification latency was the strength with which the target feature was correlated with the other features of the concept. In contrast, the major predictor of feature typicality ratings was production frequency, an individual features measure. In the model, production frequency (feature saliency) was automatically equated because a binary representation was used and all concepts were trained to an equal extent. Thus, the goal of the simulation was to demonstrate that the influence of intercorrelational strength on the activation of a target feature peaked early, then tailed off as a concept was being computed. It was assumed that the activation of a feature during the computation of a concept is monotonically related to verification latency. Furthermore, asymptotic activation might be monotonically related to feature typicality rating, although this relationship is less clear because of the greater number of intervening processes.

The model contained 14 items from Experiment 3 that differed in terms of the strength with which the target feature was correlated with other features of the concept, strong ($M = 163$, $SE = 16$), weak ($M = 37$, $SE = 9$), $t(13) = 7.60$, $p < .0001$. These groups of items were also roughly equated on three variables: production frequency, strong ($M = 12$, $SE = 2$), weak ($M = 12$, $SE = 2$), $t(13) = 0$; concept familiarity: strong ($M = 5.4$, $SE = 0.2$), weak ($M = 5.0$, $SE = 0.3$), $t(13) = 1.35$, $p > .2$; and number of individual features listed per concept: strong ($M = 18$, $SE = 1$), weak ($M = 17$, $SE = 1$), $t(13) = 1.68$, $p > .1$. In analyses restricted to these 14 items, human subjects verified strongly intercorrelated target features more quickly, $t_2(13) = 7.60$, $p < .001$, and rated them as more typical, $t_2(13) = 2.13$, $p < .05$.

For the simulation, the word form of the 28 concepts was clamped, and the concepts were allowed to converge for 10 iterations. Activation of the target feature was recorded at each iteration. Five runs with independent random starting configurations were used. A two-way, repeated-measures ANOVA was conducted with target feature activation as the dependent variable and intercorrelational strength (strong vs. weak) and iteration (1–10) as the independent variables. An interaction between intercorrelational strength and iter-

ation showed that the influence of correlated features changed over time, $F(1, 13) = 10.28$, $p < .008$. Simple main effects revealed that target features associated with concepts from the strongly intercorrelated group were not significantly more activated after Iteration 1; the first iteration was unaffected by correlated features because the activation in the conceptual units was random prior to it. Strongly intercorrelated target features were significantly more activated after Iterations 2–5 ($p < .05$). This advantage was marginal for Iterations 6–10 ($.06 < p < .1$). Intercorrelational strength was also used to predict target feature activation at each iteration. The correlation was nonsignificant for Iteration 1. Predictions were significant for Iterations 2–5, peaking at Iteration 3 ($r^2 = .37$, $p < .001$). However, the correlation between intercorrelational strength and target-feature activation tailed off and was nonsignificant for Iterations 6–10. Thus, the influence of correlated features on the time course of feature activation roughly corresponded to the results of Experiment 3.

## Discussion

A number of recent articles have been aimed at developing a theory of lexical memory based on attractor networks. This approach exemplifies general computational principles that have been outlined by McClelland (1991, 1993). Plaut, McClelland, Seidenberg, and Patterson (1996) applied them to lexical processing in stating that "processing is graded, random, adaptive, interactive, and nonlinear, and that representation and knowledge are distributed . . . these principles lead to a view in which the reading system learns gradually to be sensitive to the statistical structure among orthographic, phonological, and semantic representations, and that these representations simultaneously constrain each other in interpreting a given input" (p. 56). Evidence for this view is mounting. Plaut et al. have detailed an attractor-network account of the pronunciation of printed words and nonwords. Masson (1995) and Plaut (1995) have shown that the major empirical phenomena associated with semantic priming can be qualitatively reproduced. Hinton and Shallice (1991), Plaut and Shallice (1993), and Plaut et al. have demonstrated that attractor basins are critical to understanding the phenomena that characterize deep dyslexia. Furthermore, Kawamoto (1993) has shown that the nonlinear dynamics of attractor networks may be key to understanding lexical ambiguity resolution. The present modeling advanced the understanding of lexical memory by demonstrating that feature correlations may provide a basis for those nonlinear dynamics.

A theory of lexical memory based on attractor networks is a major departure from the traditional view. Typically, the lexicon is conceptualized as a set of nodes, each of which corresponds to a single lexical item (e.g., Morton, 1969; Coltheart, Curtis, Atkins, & Haller, 1993). Lexical access thus corresponds to activating the word's node in the orthographic or phonological lexicon. Because this node is presumed to contain a pointer to the word's meaning (i.e., meaning is directly addressable), the form of the mapping is

irrelevant. In contrast, in a distributed attractor network, the lexicon is a set of attractor basins, with each lexical item corresponding to a stable state in orthographic, phonological, and semantic space. In this system, the form of the mappings between orthography, phonology, and semantics is crucial. Indeed, the arbitrary nature of the mapping from word form to meaning has caused some researchers to suggest that it cannot be accomplished in a distributed network (e.g., Forster, 1994). However, the research presented herein demonstrates that the required structure is present in the form of correlated features.

Although most people find it easy to generalize a quasi-regular mapping (e.g., pronouncing "nust"), it is not clear whether or how generalization might proceed for a mapping that is basically arbitrary (e.g., the meaning of "nust"?). Insight might be gained by considering Plaut et al.'s (1996) analysis of how their network named nonwords. They found that nonword naming depended on subattractors that corresponded to units smaller than the patterns on which it was trained (i.e., letters, letter pairs, etc.). Furthermore, the attractor basins associated with exception words (e.g., *some*) were less componential in structure than those associated with regular words (e.g., *seed*) because correct exception pronunciation typically depends on the entire word. Thus, because the mapping from word form to meaning is generally arbitrary for monomorphemic words, little or no attractor substructure would exist to support generalization. To test the network's generalization to nonwords, the word-form representations of 10 nonwords (*abbome, zultur, keplet, freater, shamp, nector, limp, eaple, toasten, cucumder*) were clamped, and each was allowed to iterate 25 times. Few or no features were activated greater than 0.5 by these nonwords. This result coheres with the observation that English people seldom if ever generalize the word-form-to-meaning mapping. Rather, semantic generalizations typically depend on inferences based on correlated features, that is, they depend on knowledge within the semantic domain. For example, given solely the information that something ⟨has four legs⟩, people can easily provide other features that it might posses. Similarly, the model exhibited this type of generalization, performing pattern completion based on correlated features (see Appendix B). In summary, the present modeling contributed to the ongoing development of an attractor network theory of lexical representation and processing.

## GENERAL DISCUSSION

The purpose of the empirical and modeling work described above was to examine the role of featural representations in the processing of word meaning. The work addressed three general issues: the relevance of featural representations to different types of semantic tasks; the nature of featural representations, focusing on the way in which feature correlations might be learned and what their subsequent role in word recognition might be; and the organization of semantic memory, with particular emphasis on defining semantic relatedness and specifying the source of automatic semantic priming. Experiment 1 provided a

large set of feature norms that was used to construct semantic representations in terms of individual and correlated features. These representations were used to create stimuli and to predict performance in Experiments 2 and 3. Experiment 2 contrasted a speeded task in which no overt judgment of similarity was required (automatic semantic priming) with an untimed similarity-rating task. Similarity in terms of individual features predicted priming effects for artifacts but not for living things, whereas similarity in terms of correlated feature pairs predicted priming effects for living things but not for artifacts. For the untimed task, individual features predicted similarity ratings for both artifacts and living things, although the predictions were not as good for living things due to the role that knowledge of biological origin played in subjects' ratings. Correlated features did not predict similarity ratings for either type of concept. Experiment 3 also contrasted a speeded task (feature verification) with an untimed rating task (feature typicality rating), and the results corresponded to those of Experiment 2. Intercorrelational density of features, the correlated features measure, was the primary predictor of feature verification latency but did not predict feature typicality. The individual features measures (production frequency and ranked production frequency) predicted both verification latency and typicality rating.

Experiments 2 and 3 support the notion that, as in an attractor network of lexical processing, correlations among semantic features play a central role in the dynamics of computing word meaning. To illustrate this, a Hopfield (1982, 1984) network was implemented in which word form units were unidirectionally connected to semantic feature units that were fully interconnected to allow correlations to be encoded between each pair of features. The word-form-to-semantic connections functioned to put the network in a representational state within the word's semantic basin of attraction, whereas the semantic interconnections determined the rate of convergence from that point. The model simulated the priming results of Experiment 2 in that the same factors that predicted human performance also predicted the model's performance. A simulation of Experiment 3 showed that the influence of correlated features on the activation level of a target feature peaked and then tailed off during the computation of a concept. This pattern mimicked the human data in which feature correlations influenced the speeded but not the untimed task.

There were five main contributions of this work. First, the roles of featural representations and higher level knowledge in semantic tasks were integrated by carefully considering task demands. Second, a number of researchers had claimed that correlated features are a key to understanding semantic representation (e.g., Malt & Smith, 1984), but this was the first clear demonstration of their effects on tasks involving real-world concepts. Third, the model suggested a novel way of viewing how correlated features are learned and used in lexical processing. Fourth, recent studies had suggested that featural similarity was irrelevant to semantic relatedness and automatic priming (Moss et al., 1995; Shelton & Martin, 1992). However, featural similarity predicted item-by-item priming effects, suggesting that it is an impor-

tant organizing principle of semantic memory. Finally, previous research had suggested that correlated features are more prominent in the representation of living things than of artifacts (e.g., Barrett, Abdi, Murphy, & Gallagher, 1993; Gelman, 1988), and the analyses on the norms and the priming experiment provided further evidence for this claim. In addition, the network dissociated living things from artifacts on this basis without the need for qualitatively different representations.

One challenge in this research was to find appropriately yoked pairs of slow and fast tasks. Experiment 2 used two tasks that tap the similarity of lexical concepts: Automatic semantic priming reflects similarity but subjects are not required to judge it, whereas a conscious judgment is the crux of the similarity rating task. Experiment 3 used two tasks that tap the relationship between a concept and its features: Feature verification involves quickly deciding whether a feature is part of a concept, and feature-typicality rating involves specifying how typical it is of that concept. The experiments showed that these tasks differ on more than simply the speed with which they are performed; specifically, the extra time for the untimed tasks allowed for additional processes to intervene between the computation of word meaning and the decision, thus decreasing the predictive ability of correlated features. It might be possible in future research to use tasks that are even more tightly yoked than those of Experiments 2 and 3. For instance, one possibility is to compare short and long SOA conditions in a priming task. With a long SOA, subjects might generate an expectancy set to the prime and base their decisions to the target largely on this set (Neely, 1991). Although the decision would still be fast, the reliance on the expectancy set might dampen effects of correlated features. Manipulating the SOA between the concept and feature names in the verification task might show analogous effects. A third possibility might be to use short and long deadline conditions in a same–different category decision task, analogous to Goldstone's (1992) work. This task would consist of providing subjects with a number of superordinate category names prior to a block of trials. Each trial would consist of two concept names (e.g., *dog cat* or *dog chair*), and the subjects' task would be to indicate whether or not they come from the same superordinate category. The hypothesis in this case would be that the extra time allowed by the long deadline would allow for additional processing that would dampen the influence of correlated features. These and other possibilities are currently being explored in the laboratory of the first author.

## Semantic Impairments in Alzheimer's Dementia

There has been a considerable amount of research on impairments to lexical semantic representations that occur as a consequence of brain injury (e.g., stroke) or disease (e.g., herpes encephalitis; Alzheimer's disease [AD]); see Shallice (1988) for a review. One attraction of connectionist models is that they can provide a unified account of normal skilled performance and neuropsychological impairments

(see Plaut & Shallice, 1993; Plaut et al., 1996; Farah & McClelland, 1991, for examples of this approach). Our account of semantic priming effects has emphasized the importance of correlated and individual features and how their distributions differ across living-thing and artifact categories. Devlin et al. (1996) have recently examined implications of this theoretical claim about the organization of semantic memory with regard to the phenomenon of category-specific impairments. Several well-documented case studies have established that certain types of neuropathology (e.g., herpes encephalitis) sometimes result in selective impairment of biological kind or artifact categories (see Saffran & Schwartz, 1994, for review). These patterns of impairment have been explained in terms of predominant damage to a specific type of feature: perceptual in the case of biological kinds, functional in the case of artifacts (Farah & McClelland, 1991). Perceptual features are assumed to be stored in temporolimbic areas, whereas functional features are stored in frontoparietal regions. Damage localized to either region preferentially disrupts the semantic information in that region resulting in a category-specific impairment. This kind of localized damage is compatible with forms of neuropathology, such as herpes encephalitis, that are localized to specific brain regions. However, category-specific impairments have also been observed in patients with AD, a pathology of widespread, patchy damage affecting both temporolimbic and frontoparietal regions (Gonnerman, Andersen, Devlin, Kempler, & Seidenberg, in press; Silveri & Gainotti, 1988). This type of neuropathology cannot be equated with damage to particular types of units in a connectionist network. Devlin et al. have provided an account of category-specific impairments in AD using a model much like the one described earlier. The pathology associated with AD was simulated by progressively eliminating random connections between units. Devlin et al. have found biological and artifact categories behaved very differently under this type of damage. The decline in performance on items drawn from biological categories was a nonlinear function of the number of damaged connections. Biological categories withstood the effects of small amounts of damage better than artifacts; however, with additional damage, performance on biological categories degenerated in an abrupt, catastrophic manner, whereas the decrement in performance on artifacts was more gradual. Devlin et al. also have reported data from a large number of patients with AD consistent with this pattern. Patients with mild damage showed a small advantage for artifacts compared to biological kinds; with more severe degrees of impairment, biological kinds showed much more impairment than artifacts.

Devlin et al.'s (1996) account of these data implicates differences in the density of correlated features across the types of categories. The greater number of correlated features among the biological kinds initially tends to protect them from small amounts of damage. These categories can tolerate the loss of small numbers of connections because of strong correlations between features. Sufficient amounts of damage, however, result in the loss of some of these features, which contributes to impaired performance on many exemplars of a category simultaneously. In artifact catego-

ries, exemplars seem to be damaged on an item-by-item basis, as the neuropathology begins to affect features that happen to be critical to particular items but not highly correlated with other features.

Devlin et al.'s (1996) research suggests that the theoretical analysis of featural representations of meaning that we have proposed can also account for some phenomena concerning the time course of the effects of progressive dementing disease. The explanatory value of a theoretical approach greatly increases when the same principles (in this case, about lexical representation) can be used to explain very different kinds of behavioral phenomena. A second point is that these assumptions about lexical representation are compatible with effects of different types of neuropathology. Damaging a bank of units is different from progressively eliminating connections between units. The former provides a way to investigate effects of localized brain damage, the latter the diffuse effects of AD. Devlin et al. showed that these types of damage give rise to different profiles of impaired performance that correspond well to the behaviors seen in different types of patients. Although this research only addresses a narrow range of behavioral impairments, it suggests that it should be possible to use models of normal performance to account for effects of different types of neuropathology.

## Correlations Versus Relations

In Experiments 2 and 3, knowledge of correlated features influenced early processing, whereas later processing was dominated by individual features. Paradoxically, Goldstone (1992) and Ratcliff and McKoon (1989) have found that subjects base decisions on individual features when they are hurried, whereas they use relational information only when given sufficient time to process the stimuli. This inconsistency is easily resolved, however, by contrasting the type of information used in each case. For example, in Goldstone's experiments, subjects judged the similarity of visually presented scenes. Relational information was used to derive a common orientation between scenes before performing a feature-by-feature comparison. Thus, the relational information was novel and had to be computed anew on-line. On the other hand, in the present research, no new relational information was required to perform the priming or verification tasks. Unlike the studies of Goldstone and McKoon and Ratcliff, the correlational knowledge consisted of established statistical information about the distribution of features of common lexical concepts. The fact that it influenced early processing can be taken as evidence for the view that this knowledge is inherent to the mechanism that computes word meaning.

## Conceptual Development

The attractor network described earlier continues a tradition of theories and models that have emphasized the importance of capturing predictive structure in the environment as a primary means of learning concepts and

categories (Anderson, 1991; Brunswick, 1955; Clapper & Bower, 1991; Rosch, 1978). For example, Billman and Heit (1988) have described a model that is similar to ours in which the learning of feature correlations is a central process of conceptual development. Furthermore, their models stress that these correlations are learned from observation. Their models differ, however, in that they deal with explicit learning of feature correlations; they are framed in terms of learning correlational rules. In contrast, using an attractor network as a metaphor for learning implies that much of a person's knowledge of correlated features is of an implicit nature; children automatically learn feature correlations that then directly influence conceptual processing.

The explicit–implicit dimension is relevant to treatments of correlated features and has played a role throughout this article. Recent claims made by Holyoak and Spellman (1993) concerning implicit learning imply that a considerable amount of people's knowledge of correlated features may be implicit. They claimed that implicit knowledge often takes the form of covariations in the environment and that people learn these covariations by exposure to stimuli exhibiting them, often without intention or awareness. Holyoak and Spellman also noted that connectionist models are particularly well suited for encoding and using this type of knowledge. Strongly opposing this view, however, are Medin, Wattenmaker, and Hampson (1987) and Murphy and Wisniewski (1989), who have claimed that a correlation between two features cannot be learned unless a prior theoretical relation exists between them. The human data are somewhat equivocal on this matter. Experiments such as Billman (1989), Billman and Knutson (1996), and Waldmann et al. (1995) have found evidence in adults for statistical learning of feature correlations without prior theories linking the features. In addition, Younger and Cohen (1983) have shown that 10-month-old infants are sensitive to the distributional properties of features of novel animals. However, failures have also been reported, such as Murphy and Wisniewski (1989) and Wattenmaker (1993). One factor that may have led to these failures is the lack of training time. Typically, in studies that have failed to find effects of correlated features, subjects are exposed to a relatively small number of novel stimuli for short durations. Thus, null effects may have resulted simply because subjects were not provided with sufficient opportunity to encode the requisite bottom-up knowledge. Furthermore, these studies often involve novel combinations of familiar features. If it is assumed that subjects enter these experiments possessing a great deal of knowledge about feature correlations, it is unclear how much training would be required to overturn this prior learning and rearrange their knowledge base so that it mirrored the distributional patterns in the experimental training set.

Finally, we believe that theories do play a role in concept learning or the learning of correlated features. It is easy to imagine that statistical knowledge and relational theories mutually facilitate conceptual development. On the one hand, although there may be a great deal of statistical knowledge of feature co-occurrence that is rarely brought to consciousness (Holyoak & Spellman, 1993), people are

capable of doing so. For example, it is probable that few people consciously note the statistical correlation between ⟨has fur⟩ and ⟨has a tail⟩. However, if it was brought to their attention, or they did consciously note it for whatever reason, it could trigger a search for, or form the basis of, a theory. On the other hand, being taught certain facts or theories might lead to increased attention to specific features or pairs of features, which would lead in turn to enhanced encoding of a statistical regularity through some mechanism such as focused sampling. Along these lines, Barrett et al. (1993) showed that a child's sensitivity to correlated features was modulated by whether the relationship made sense. Unfortunately, the learning aspect of our model did not include this interplay between theoretical and statistical knowledge, and it stands as a challenge for connectionists and symbolic modelers to extend models in this way. However, recent research has begun to deal with the relationships between statistical and theory-based knowledge in conceptual development (Wisniewski, 1995; Wisniewski & Medin, 1994), and this should continue to be an important topic of future research.

## References

Anderson, J. R. (1991). The adaptive nature of human categorization. *Psychological Review, 98,* 409–429.

Ashcraft, M. H. (1978). Feature dominance and typicality effects in feature statement verification. *Journal of Verbal Learning and Verbal Behavior, 17,* 155–164.

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

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

Barsalou, L. W. (1982). Context-independent and context-dependent information in concepts. *Memory & Cognition, 10,* 82–93.

Barsalou, L. W. (1987). The instability of graded structure: Implications for the nature of concepts. In U. Neisser (Ed.), *Concepts and conceptual development—ecological and intellectual factors in categorization: Emory Symposia in Cognition* (Vol. 1, pp. 101–140). Cambridge, England: Cambridge University Press.

Barsalou, L. W. (1989). Intraconcept similarity and its implications for interconcept similarity. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 76–121). Cambridge, England: Cambridge University Press.

Barsalou, L. W., Olseth, K. L., & Wu, L. (1996). *Producing features for concepts.* Manuscript in preparation.

Battig, W. F., & Montague, W. E. (1969). Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology Monographs, 80*(3, Pt. 2).

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

Billman, D. (1989). Systems of correlations in rule and category learning: Use of structured input in learning syntactic categories. *Language and Cognitive Processes, 4,* 127–155.

Billman, D., & Heit, E. (1988). Observational learning from in-

ternal feedback: A simulation of an adaptive learning method. *Cognitive Science, 12,* 587–625.

Billman, D., & Knutson, J. (1996). Unsupervised concept learning and value systematicity: A complex whole aids learning the parts. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 458–475.

Brunswick, E. (1955). Symposium on the probability approach in Psychology: Representative design and probabilistic theory in a functional psychology. *Psychological Review, 12,* 103–217.

Clapper, J. P., & Bower, G. H. (1991). Learning and applying category knowledge in unsupervised domains. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 27, pp. 65–108). Toronto, Ontario, Canada: Academic Press.

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

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

Coltheart, M., Curtis, B., Atkins, P., & Haller, M. (1993). Models of reading aloud: Dual-route and parallel-distributed-processing approaches. *Psychological Review, 100,* 589–608.

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

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

Devlin, J. T., Gonnerman, L. M., Andersen, E. S., & Seidenberg, M. S. (1996). *Category specific semantic deficits in focal and widespread brain damage: A computational account.* Manuscript submitted for publication.

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

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

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

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

Gelman, S. A., & Wellman, H. M. (1991). Insides and essences: Early understanding of the non-obvious. *Cognition, 38,* 213–244.

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science, 7,* 155–170.

Gluck, M. A., Bower, G. H., & Hee, M. R. (1989). A configural-cue network model of animal and human associative learning. *Proceedings of the Eleventh Annual Conference of the Cognitive Science Society, 11,* 323–332.

Goldstone, R. L. (1992). Locally to globally consistent processing in similarity. *Proceedings of the Fourteenth Annual Conference of the Cognitive Science Society, 14,* 337–342.

Gonnerman, L. M., Andersen, E. S., Devlin, J. T., Kempler, D., & Seidenberg, M. S. (in press). Double dissociation of semantic categories in Alzheimer's disease. *Brain and Language.*

Gough, P. B., & Cosky, M. J. (1977). One second of reading again.

In N. J. Castellan, Jr., D. B. Pisoni, & G. R. Potts (Eds.), *Cognitive theory* (Vol. 2, pp. 271–288). Hillsdale, NJ: Erlbaum.

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

Hertz, J., Krogh, A., & Palmer, R. G. (1991). *Introduction to the theory of neural computation: Santa Fe Institute Studies in the Sciences of Complexity lecture notes* (Vol. 1). Workingham, England: Addison-Wesley.

Hinton, G. E. (1981). Implementing semantic networks in parallel hardware. In G. E. Hinton & J. A. Anderson (Eds.), *Parallel models of associative memory* (pp. 161–188). Hillsdale, NJ: Erlbaum.

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

Hintzman, D. L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review, 93,* 411–428.

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

Holyoak, K. J., & Spellman, B. A. (1993). Thinking. *Annual Review of Psychology, 44,* 265–315.

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

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

Humphrey, G. K., Goodale, M. A., Jacobson, L. S., & Servos, P. (1994). The role of surface information in object recognition: Studies of a visual form agnosic and normal subjects. *Perception, 23,* 1457–1481.

Jared, D., & Seidenberg, M. S. (1992). Does word identification proceed from spelling to sound to meaning? *Journal of Experimental Psychology: General, 120,* 358–394.

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

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

Keil, F. C. (1989). *Concepts, kinds, and cognitive development.* Cambridge, MA: MIT Press.

Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review, 99,* 22–44.

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

Malt, B. C., & Johnson, E. C. (1992). Do artifact concepts have cores? *Journal of Memory and Language, 31,* 195–217.

Malt, B. C., & Smith, E. E. (1984). Correlated features in natural categories. *Journal of Verbal Learning and Verbal Behavior, 23,* 250–269.

Markman, A. B., & Gentner, D. (1993). Splitting the differences: A structural alignment view of similarity. *Journal of Memory and Language, 32,* 517–535.

Masson, M. E. J. (1995). A distributed memory model of semantic priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 1155–1172.

McClelland, J. L. (1991). Stochastic interactive processes and the effect of context on perception. *Cognitive Psychology, 23,* 1–44.

McClelland, J. L. (1993). The GRAIN model: A framework for modeling the dynamics of information processing. In D. E.

Meyer & S. Kornblum (Eds.), *Attention and Performance 14: Synergies in experimental psychology, artificial intelligence, and cognitive neuroscience* (pp. 655–688). Cambridge, MA: MIT Press.

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

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

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

McRae, K. (1991). *Independent and correlated properties in artifact and natural kind concepts*. Unpublished doctoral dissertation, McGill University, Montreal, Quebec, Canada.

McRae, K., & Boisvert, S. (1997). *The importance of automatic semantic relatedness priming for distributed models of word meaning*. Manuscript submitted for publication.

McRae, K., Jared, D., & Seidenberg, M. S. (1990). On the roles of frequency and lexical access in word naming. *Journal of Memory and Language, 29*, 43–65.

Medin, D. L. (1989). Concepts and conceptual structure. *American Psychologist, 44*, 1469–1481.

Medin, D. L., Altom, M. W., Edelson, S. M., & Freko, D. (1982). Correlated symptoms and simulated medical classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 8*, 37–50.

Medin, D. L., Goldstone, R. L., & Gentner, D. (1993). Respects for similarity. *Psychological Review, 100*, 254–278.

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review, 85*, 207–238.

Medin, D. L., & Shoben, E. J. (1988). Context and structure in conceptual combination. *Cognitive Psychology, 20*, 158–190.

Medin, D. L., Wattenmaker, W. D., & Hampson, S. E. (1987). Family resemblance, conceptual cohesiveness, and category construction. *Cognitive Psychology, 19*, 242–279.

Minsky, M. (1975). A framework for representing knowledge. In P. H. Winston (Ed.), *The psychology of computer vision* (pp. 211–277). New York: McGraw-Hill.

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

Moscovitch, M., Goshen-Gottstein, Y., & Vriezen, E. (1994). Memory without conscious recollection: A tutorial review from a neuropsychological perspective. In C. Umilta & M. Moscovitch (Eds.), *Attention and Performance 15: Conscious and nonconscious information processing.* (pp. 619–660). Cambridge, MA: MIT Press.

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., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review, 92*, 289–316.

Murphy, G. L., & Wisniewski, E. J. (1989). Feature correlations in conceptual representations. In G. Tiberghien (Ed.), *Advances in cognitive science: Vol. 2. Theory and applications* (pp. 23–45). Chichester, England: Ellis Horwood.

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

Nebes, R. N., Brady, C. B., & Huff, F. J. (1989). Automatic and attentional mechanisms of semantic priming in Alzheimer's disease. *Journal of Clinical and Experimental Neuropsychology, 11*, 219–230.

Neely, J. H. (1977). Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited-capacity attention. *Journal of Experimental Psychology: General, 106*, 22–254.

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

Neely, J. H., & Keefe, D. E. (1989). Semantic context effects on visual word processing: A hybrid prospective/retrospective processing theory. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 24, pp. 207–248). New York: Academic Press.

Neuman, P. G. (1974). An attribute frequency model for the abstraction of prototypes. *Memory & Cognition, 2*, 241–248.

Norman, D. A., & Rumelhart, D. E. (1975). *Explorations in cognition*. San Francisco: Freeman.

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

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

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

Posner, M. I., & Snyder, C. R. R. (1975). Facilitation and inhibition in the processing of signals. In P. M. A. Rabbitt & S. Dornic (Eds.), *Attention and Performance 5* (pp. 669–682). New York: Academic Press.

Pylyshyn, Z. (1984). *Computation and cognition: Toward a foundation for cognitive science*. Cambridge, MA: MIT Press.

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

Ratcliff, R., & McKoon, G. (1989). Similarity information versus relational information: Differences in the time course of retrieval. *Cognitive Psychology, 21*, 139–155.

Rayner, K. (1978). Eye movements in reading and information processing. *Psychological Bulletin, 85*, 618–660.

Rips, L. J. (1989). Similarity, typicality, and categorization. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 21–59). Cambridge, England: Cambridge University Press.

Roediger, H. L., & McDermott, K. B. (1993). Implicit memory in normal human subjects. In H. Spinnler & F. Boller (Eds.), *Handbook of Neuropsychology* (Vol. 8, pp. 63–130). Amsterdam: Elsevier.

Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General, 104*, 192–233.

Rosch, E. (1978). Principles of categorization. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization* (pp. 27–48). Hillsdale, 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, 8*, 382–439.

Saffran, E. M., & Schwartz, M. F. (1994). Of cabbages and things: Semantic memory from a neuropsychological perspective—A

tutorial review. In C. Umilta & M. Moscovitch (Eds.), *Attention and Performance 15: Conscious and nonconscious information processing* (pp. 507–536). Cambridge, MA: MIT Press.

Schacter, D. L. (1992). Priming and multiple memory systems: Perceptual mechanisms of implicit memory. *Journal of Cognitive Neuroscience, 4,* 244–256.

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

Shallice, T. (1988). *From neuropsychology to mental structure.* Cambridge, England: Cambridge University Press.

Sharkey, N. E. (1989). The lexical distance model and word priming. *Proceedings of the Eleventh Annual Conference of the Cognitive Science Society, 11,* 860–867.

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

Shepard, R. (1958). Stimulus and response generalization: Deduction of the generalization gradient from a trace model. *Psychological Review, 65,* 242–256.

Shepard, R. (1987). Toward a universal law of generalization for psychological space. *Science, 237,* 1317–1323.

Shimamura, A. P., & Squire, L. R. (1984). Paired-associate learning and priming effects in amnesia: A neuropsychological study. *Journal of Experimental Psychology: General, 113,* 556–570.

Silveri, M. C., & Gainotti, G. (1988). Interaction between vision and language in category-specific semantic impairments. *Cognitive Neuropsychology, 5,* 677–709.

Smith, E. E., & Medin, D. L. (1981). *Categories and concepts.* Cambridge, MA: Harvard University Press.

Smith, E. E., & Osherson, D. N. (1984). Conceptual combination with prototype concepts. *Cognitive Science, 8,* 337–361.

Smith, E. E., Osherson, D. N., Rips, L. J., & Keane, M. (1988). Combining prototypes: A selective modification model. *Cognitive Science, 12,* 485–527.

Tsodyks, M. V., & Feigelman, M. V. (1988). The enhanced storage capacity in neural networks with low activity level. *Europhysics Letters, 6,* 101–105.

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

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185,* 1124–1131.

Waldmann, M. R., Holyoak, K. J., & Fratianne, A. (1995). Causal models and the acquisition of category structure. *Journal of Experimental Psychology: General, 124,* 181–206.

Wattenmaker, W. D. (1991). Learning modes, feature correlations, and memory-based categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 203–222.

Wattenmaker, W. D. (1993). Incidental concept learning, feature frequency, and correlated properties. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17,* 908–923.

Whitten, W. B., II, Suter, W. N., & Frank, M. L. (1979). Bidirectional synonym ratings of 464 noun pairs. *Journal of Verbal Learning and Verbal Behavior, 18,* 109–127.

Wisniewski, E. J. (1995). Prior knowledge and functionally relevant features in concept learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 449–468.

Wisniewski, E. J., & Medin, D. L. (1994). On the interaction of theory and data in concept learning. *Cognitive Science, 18,* 221–281.

Younger, B. A., & Cohen, L. B. (1983). Infant perception of correlations among attributes. *Child Development, 54,* 858–867.

# Appendix A

## The Stimuli of Experiments 1 and 2

Prime-target pairs are grouped together on a line (e.g., bird–robin, eagle–hawk, duck–chicken, etc.). An asterisk (*) indicates pairs included in the model.

### Prime-Target Pairs

| | | | | | |
|---|---|---|---|---|---|
| bird | robin | eagle* | hawk* | duck* | chicken* |
| emu | ostrich | vulture* | buzzard* | budgie* | parakeet* |
| stork | crane | starling | crow | canary* | finch* |
| goose | turkey | | | | |
| mammal | pig | caribou | moose | tiger* | lion* |
| goat | sheep | dog* | cat* | rat* | mouse* |
| whale | dolphin | horse* | pony* | cow* | bull* |
| deer | fawn | | | | |
| fruit | apple | orange | grapefruit | lemon* | lime* |
| cherry | cranberry | cantaloupe | honeydew | peach* | nectarine* |
| prune | plum | coconut* | pineapple* | raisin* | grape* |
| mandarin | tangerine | | | | |
| vegetable | corn | garlic | onions | cucumber* | zucchini* |
| broccoli | cauliflower | squash | pumpkin | carrot* | celery* |
| peas | beans | potato* | yam* | radish* | beets* |
| lettuce | cabbage | | | | |
| clothing | socks | nylons | leotards | shoes* | boots* |
| sandals | slippers | mittens | gloves | skirt* | dress* |
| tie | belt | pants* | trousers* | shirt* | blouse* |
| camisole | bra | | | | |
| furniture | chair | closet | dresser | couch* | sofa* |
| carpet | mat | shelves | cupboard | drapes* | curtains* |
| bureau | desk | cushion | pillow | lamp* | chandelier* |
| stereo | radio | | | | |
| kitchen | stove | mug | cup | saucer* | plate* |
| spoon | fork | faucet | sink | pot* | pan* |
| mixer | blender | microwave* | toaster* | fridge* | freezer* |
| bottle | jar | | | | |
| tool | hammer | screws | bolts | wrench* | pliers* |
| shed | barn | level | ruler | hoe* | shovel* |
| file | sandpaper | crayon* | pencil* | drill* | screwdriver* |
| vice | clamp | | | | |
| vehicle | car | tricycle | bike | truck* | van* |
| subway | bus | canoe | raft | scooter* | motorcycle* |
| wagon | cart | jet* | airplane* | ship* | yacht* |
| dunebuggy | jeep | | | | |
| weapon | gun | slingshot | catapult | pistol* | rifle* |
| dagger | knife | rock | stone | cannon* | bazooka* |
| axe | tomahawk | spear* | sword* | missile* | bomb* |
| club | stick | | | | |

*(Appendixes continue)*

## Appendix B

## Description of the Model

### Determining Connection Weights

The weights between semantic units were determined independently of the weights connecting the word form to the semantic units. Training for each component was accomplished using a single-batch matrix multiplication rather than iteratively training on each pattern. That is, the network was trained in a single step rather than presenting patterns to the network one at a time until some learning criterion was met. Mathematically, the methods are similar, with matrix multiplication being computationally preferable because it is faster. However, a model trained in this way does not provide information about the time course of learning.

All concepts were equally familiar to the model because each was used once when the weights were determined. The semantic ⟺ semantic weights were set using the Hopfield (1982, 1984) learning rule, slightly modified to be optimal for sparse patterns (Tsodyks & Feigelman, 1988). The learning rule was (using the symbols from Figure 3)

$$w_{jk} = 1/n_s \sum_p [(x_{jp} - m_p)(x_{kp} - m_p)], \qquad (B1)$$

where $w_{jk}$ represented the weight of the connection between $unit_j$ and $unit_k$ (connections were symmetric, $w_{jk} = w_{kj}$), $n_s$ was the number of semantic units (646), and $x_{jp}$ represented the presence (1) or absence (0) of $feature_j$ in $concept_p$. The term $m_p$ was the number of features possessed by $concept_p$ expressed as a proportion of the total number of features, that is, the number of features in $concept_p$ divided by 646. Including this term has two advantages. First, it increases storage capacity (Tsodyks & Feigelman). This is important because we wanted to train the network on as many patterns as possible to give it exposure to features that co-occur across concepts. Second, as with the Pearson product measure $(r)$ used for the correlated features representation, this learning rule is sensitive to the fact that concepts are extremely sparse in semantic space as defined by individual features. For example, in the model, *eagle* possessed 17 of 646 features, a proportion of .026. If two *eagle* features were present (e.g., ⟨has wings⟩ and ⟨flies⟩), their connection was strengthened by $(1 - .026)(1 - .026) = 0.948$. If two features were absent (e.g., ⟨made of wood⟩ and ⟨has a handle⟩), their connection was strengthened by $(0 - .026)(0 - .026) = 0.0007$. Finally, the connection between a present and an absent feature (e.g., ⟨has wings⟩ and ⟨has a handle⟩) was strengthened in the negative direction by an intermediate quantity, $(1 - .026)(0 - .026) = -0.0256$. This learning rule captures the following intuition. Because a concept possesses so few features in relation to the vast number of possible features in the world, the absence of a feature generally carries little or no information. For example, when someone attends to a dog, she notices the features it possesses, but does not tend to notice ones it does not possess, such as ⟨has a handle⟩ or ⟨has wheels⟩. Tsodyks and Feigelman's learning rule correctly treats simultaneously present features as more important than simultaneously absent ones. Finally, it should be noted that the situation in which a dog is missing a typical *dog* feature is qualitatively different; if a dog does not have four legs, people do tend to notice (and so would the model, see below).

The word form ⟹ semantic connections were computed using a similar Hebbian correlational learning rule (Hebb, 1949).

$$z_{ij} = 1/n_{wf} \sum_p [(a_{ip})(x_{jp} - m_p)] \qquad (B2)$$

The variable $a_{ip}$ was the normalized word-form representation for $unit_i$; normalization removed effects of word length. The variables $x_{jp}$ and $m_p$ were the same as in Equation 1, and $n_{wf}$ was the number of word-form units (379). Note that there were no $z_{ji}$ connections. A word-form pattern was represented by a 1 at each unit corresponding to a letter triple contained in the word, and zeroes elsewhere. Patterns were normalized to remove word-length bias according to the following equation:

$$\Sigma a_{ip}^2 = 4 \quad \text{for each } concept_p. \qquad (B3)$$

### Computing a Word's Featural Representation

The dynamics of unit activation were determined by

$$x_i(t + 1) = g(c_1 \sum_j (z_{ji} \times a_i) + c_2 \sum_j (w_{ji} \times x_j(t)) - \emptyset), \qquad (B4)$$

where $g(y) = 0.5 \tanh(c_3 y) + 0.5$.

The variable $x_i(t)$ was the activation of $unit_i$ at time $t$. The variables $a$, $z$, and $w$ were defined earlier. Synchronous updating was used for speed and ease of simulation, and there is no reason to expect qualitative differences with asynchronous updating. The following values were used for the simulations presented: $c_1 = 0.85$, $c_2 = 0.33$, $\emptyset = 0.0105$, $c_3 = 400$, and $g$ was a steep sigmoid; similar results were obtained with various combinations of values for the constants. The algorithm was not overly sensitive to the steepness of the sigmoid (given by $c_3$), but a step function did not work (see next paragraph).

To compute a concept, its word-form representation was used as input. Thus, it was assumed that preliminary visual or auditory analyses had resulted in an internalized distributed representation of the word's spelling or pronunciation. The network was initialized by setting 60 randomly chosen semantic units to .25. Because concepts activated an average of approximately 15 units, when 60 units were set to .25, the total activation in the system was similar to its normal stable state. Given that activation began as random and close to zero, in the first iteration, activation of semantic units resulted primarily from word form and put the network into the correct basin of attraction. Further iterations resulted in a descent in semantic state space toward the lowest point of the basin, which corresponded to the learned concept. Thus, the sigmoidal activation function performed better than the step function because it allowed semantic unit activation to accrue slowly, thereby increasing the degree of mutual facilitation among units.

In summary, the model can be understood as a Hopfield (1982, 1984) network with the word form ⟹ semantic connections acting as thresholds that were variable over units and patterns, and the semantic interconnections determining the general topology of the energy function, that is, which combinations of features were stable based on its experience with feature co-occurrence. Thus, the model embodied two important principles: It naturally learned how features co-occurred in the concepts on which it was trained, and it used this knowledge to drive the system to a stable state.

## Performance Analyses

The model successfully learned 82 of the 84 concepts on which it was trained; this was approximately the maximum number of concepts that could be learned by the network due to cross talk among concept vectors. After training, the network's retention of each concept was tested by randomly initializing 60 semantic units to a value of .25, then clamping the word-form representation and allowing the model to iterate ten times. Two performance measures were recorded after each iteration: (a) error, the sum of squared error between the computed concept vector and the target vector; and (b) activated features, where a unit was counted as "on" if its activation was greater than .5, and "off" otherwise. Results were averaged over five runs, each of which used a new set of randomly chosen initial values for the semantic units.

Figure B1 shows a graph of the mean sum of squared error for the 84 concepts, as well as for *crayon*, the most slowly converging concept, and *jet*, a concept that converged, then activated a number of incorrect features. On average, error steadily decreased until Iteration 5, at which time it stabilized. That is, concepts tended to stabilize after five iterations, at which point the mean error was 0.623. This error was quite low; by comparison, if the network had simply turned all units off, which might have been a reasonable solution given the sparsity of the patterns, the mean expected error would have been 16.2. The concepts *crayon* and *jet* were interesting and contrasting cases. Because the features of *crayon* were sparsely intercorrelated, it was slow to converge. The concept *jet*, on the other hand, contained features that were densely intercorrelated, so that its semantic representation converged quickly. However, the features of *jet* were also highly correlated with a number of bird-like features so that although it converged quickly, error subsequently increased when some of them became activated (⟨isa bird⟩, ⟨has feathers⟩, ⟨has a beak⟩, and ⟨eats⟩).

In general, errors made by the network followed a similar pattern to *jet*. After 4 iterations, 10 features from 8 concepts were activated that were not part of those concepts according to the representations from Experiment 1 (i.e., how the network was trained). By 7 iterations, it had increased to 29 features from 16 concepts, and was stable through 10 iterations. The 16 concepts for which additional features were activated contained features that were strongly correlated with one another. A measure called intercorrelational density was calculated as the sum of the percentages of variance shared by significantly correlated feature pairs within a concept; the mean intercorrelational density of the 16 concepts (979) was much higher than the average of all 84 (579). The most interesting aspect of these "errors" of inclusion was that, for 14 of the 16 concepts, the model activated valid features, even though less than 5 of 30 subjects had listed those features in the norms. For example, according to the model, a *budgie* ⟨has wings⟩, a *buzzard* and an *eagle* ⟨are animals⟩, a *chicken* and a *duck* ⟨are
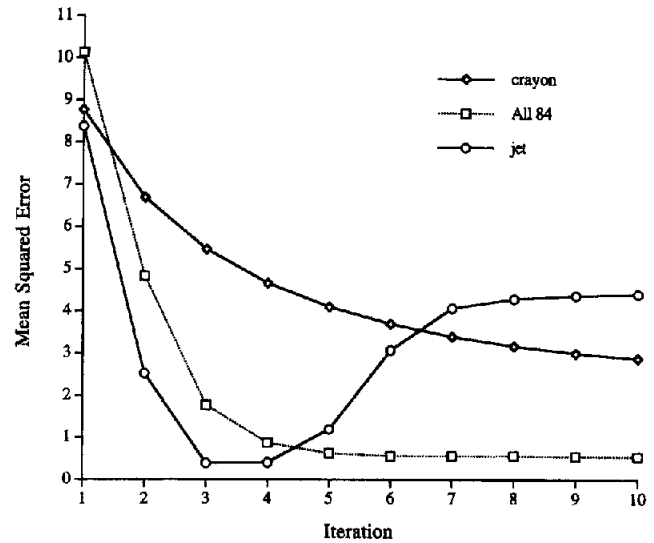


*Figure B1.* Convergence patterns for the average of the 84 concepts, as well as *crayon*, the most slowly converging concept, and *jet*, which converged then activated additional features.

large⟩, a *cannon* ⟨is dangerous⟩, a *carrot*, a *radish*, and a *zucchini* ⟨are edible⟩, a *carrot* ⟨has leaves⟩, a *hawk* ⟨eats⟩, a *mouse* ⟨has four legs⟩, a *missile* ⟨is loud⟩, and *trousers* ⟨are worn by women⟩. However, although the model correctly believes that a *canary* is ⟨an animal⟩, it also thinks that it ⟨is loud⟩ and ⟨is large⟩. As well, for both *jet* and *cat*, birdlike features were erroneously activated.

In contrast, 12 concepts whose features were sparsely intercorrelated (mean intercorrelational density = 135) were slow to converge because mutual activation among each concept's features was relatively weak. After 4 iterations, the activation of 62 features distributed across these 12 concepts remained erroneously below .5. After 7 iterations, 10 features of *crayon* remained activated below .5, as well as one feature of *yam*. After 10 iterations, only *yam*—⟨is like a potato⟩ remained incorrectly off.

In summary, the model learned 82 of the 84 concepts to the point where, after 7–10 iterations, only a few feature units were activated that were not part of the representation as specified by the norms. Furthermore, these additional feature units were activated because they were highly correlated with other features within the concept and were, for the most part, appropriate. This process of filling in additional information about a concept can be considered a form of inference and illustrates that the network is a pattern-completion device that relies on its knowledge of feature correlations.

## Appendix C

### The Stimuli of Experiment 3: Strongly and Weakly Intercorrelated Groups

| Feature | Strong | Weak |
|---|---|---|
| come (comes) in different colours | pants | carpet |
| eats | parakeet | whale |
| found in bathrooms | faucet | mat |
| found in kitchens | microwave | faucet |
| has an engine | motorcycle | yacht |
| has a handle | tomahawk | wagon |
| has a long tail | mouse | pony |
| has a seat | tricycle | chair |
| has leaves | lettuce | pineapple |
| has legs | caribou | canary |
| has skin | apple | potato |
| has teeth | lion | rat |
| has wheels | scooter | cannon |
| hunted by people | deer | ducks |
| is (are) comfortable | couch | slippers |
| is (are) decorative | drapes | carpet |
| is (are) transparent | jar | nylons |
| is breakable | plate | crayon |
| is colourful | budgie | carpet |
| is crunchy | carrot | apple |
| is dangerous | rifle | motorcycle |
| is electrical | blender | drill |
| is fast | car | microwave |
| is green | lettuce | lime |
| is round | peach | cabbage |
| is soft | sofa | carpet |
| is sweet | peach | raisin |
| is tropical | coconut | parakeet |
| made of material | pants | couch |
| made of plastic | jar | sandals |
| migrates | goose | caribou |
| used as a pet | parakeet | dog |
| used as a weapon | axe | stick |
| used for protection | pistol | dog |
| used for transportation | van | raft |
| used in the circus | lion | cannon |
| worn for warmth | gloves | shirt |