Categorical structure among shared features in networks of early-learned nouns

Thomas T. Hills [a,*], Mounir Maouene [b], Josita Maouene [c], Adam Sheya [c], Linda Smith [c]

[a] Cognitive and Decision Sciences, University of Basel, 4056 Basel, Switzerland
[b] UFR: Artificial Intelligence and Bioinformatics, ENSA, Tangier, Morocco
[c] Department of Psychological and Brain Sciences, Indiana University, USA

ARTICLE INFO

ABSTRACT

The shared features that characterize the noun categories that young children learn first are a formative basis of the human category system. To investigate the potential categorical information contained in the features of early-learned nouns, we examine the graph-theoretic properties of noun–feature networks. The networks are built from the overlap of words normatively acquired by children prior to 2½ years of age and perceptual and conceptual (functional) features acquired from adult feature generation norms. The resulting networks have small-world structure, indicative of a high degree of feature overlap in local clusters. However, perceptual features – due to their abundance and redundancy – generate networks more robust to feature omissions, while conceptual features are more discriminating and, per feature, offer more categorical information than perceptual features. Using a network specific cluster identification algorithm (the clique percolation method) we also show that shared features among these early-learned nouns create higher-order groupings common to adult taxonomic designations. Again, perceptual and conceptual features play distinct roles among different categories, typically with perceptual features being more inclusive and conceptual features being more exclusive of category memberships. The results offer new and testable hypotheses about the role of shared features in human category knowledge.

© 2009 Elsevier B.V. All rights reserved.

1. Introduction

Theories about categories are often about shared features and how lower-order categories can be organized into higher-order categories by their overlapping feature distributions (McRae, Cree, Westmacott, & De Sa, 1999; Rogers & McClelland, 2004; Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976). Although the relevance of shared features to category formation is generally well accepted, there are theoretical disputes about whether shared features in and of themselves are sufficient to form meaningful categories and also as to whether some kinds of features are more important than others (Ahn, Kalish, Medin, & Gelman, 1995; De Renzi & Lucchelli, 1994;

Komatsu, 1992). With respect to the question of superordinate category formation, the developmental literature has been particularly concerned with so-called perceptual versus conceptual features (Booth & Waxman, 2002; Gelman & Bloom, 2000; Keil, 1979, 1989; Kemler Nelson, Russell, Duke, & Jones, 2000; Madole & Oakes, 1999; Mandler & McDonough, 1996; Nelson & Ware, 2002; Quinn & Eimas, 1996; Smith, Jones, & Landau, 1996). This paper takes a first look at the shared-feature structure of categories commonly known to children younger than 3 years of age, using a graph-theoretic approach to understand how shared features in general, and perceptual and conceptual features in particular, may contribute to early category knowledge.

In the literature on children's categories, "perceptual features" refer to the perceivable and fixed properties of an individual thing (e.g., "has wheels"). In contrast,

---

* Corresponding author. Tel.: +41 (0) 61 267 3521.
  E-mail address: thomas.hills@unibas.ch (T.T. Hills).

"conceptual features" concern relations (also perceivable) that do not so much characterize an individual thing as its role in some event (e.g., "used for transportation"). One controversy concerns whether perceptual features are the developmentally earlier source of children's category organizations (with conceptual features emerging later) or whether more relational and conceptual features organize categories from the start. This debate is also related to proposals that conceptual features are privileged in superordinate category formation (Carey, 1985; Gelman, 1990; Mandler, 1992a, 1992b) and in licensing causal inferences about different kinds (Keil, 1994; Waxman & Markow, 1995; Younger & Cohen, 1990).

There is considerable evidence on both sides, including studies showing that infants and young children readily learn about correlated perceptual features (e.g., Mareschal, Quinn, & French, 2002; Quinn & Eimas, 1996; Rakison, 2003, 2005; Younger & Cohen, 1990) and often have difficulty using conceptual features (Carey, 1985; Keil & Batterman, 1984; Landau, Smith, & Jones, 1988; Sheya & Smith, 2006); while other studies show that relational features support category inferences by young children and often trump perceptual features (e.g., Gelman & Bloom, 2000; Graham & Kilbreath, 2007; Kemler Nelson, Frankenfield, Morris, & Blair, 2000). The details of this debate are not addressed in this paper. Instead, as Sheya and Smith (2006) argue, the evidence as a whole clearly indicates that both kinds of features matter. What is needed, then, is a better understanding of their inter-related roles in a larger system of developing categories.

One way to approach this question is by examining the graph-theoretic properties of early noun–feature networks, with an eye towards the distinctive contributions made by different feature types. The idea that semantic knowledge may be understood as networks of interconnected concepts has been around since the beginning of cognitive science (e.g., Quilian, 1967; Rumelhart & Norman, 1973; Shapiro, 1976). Many have claimed that categorical knowledge can be derived from the structure of these representations – for example, from the way features are correlated across nouns (Rosch et al., 1976; Rogers & McClelland, 2004). We take the feature correlation approach to build network representations, and then use the formalisms of graph theory to examine the noun–feature relationships in terms of the structure they provide.

One example of the graph-theoretic approach is the evaluation of small-world structure in large-scale semantic networks (Steyvers & Tenenbaum, 2005; Vitevitch, 2008). A small-world network is a network in which the local clustering among nodes is high, despite the fact that the average distance between any two nodes is not dramatically different from what one would expect from a random network with the same density – i.e., having the same number of nodes and links (Dorogotsev & Mendes, 2003). Small-world structure seems a likely characteristic of feature-based categories for two reasons: (1) nouns need to belong to local clusters of items that are conceptually similar (i.e., categories), but may be sufficiently discriminated from sharing connections with random nouns, and (2) some nouns belong to multiple clusters (e.g., AIRPLANE is a flying thing and a vehicle). By examining the small-world

structure of early noun networks – as well as other graph-theoretic properties – we take a quantitative approach to evaluating the structural contributions of perceptual and conceptual features in the development of early categories.

Accordingly, this study examines the graph-theoretic properties of the system of pairwise relations among early-learned noun categories as indicated by the shared features that connect them. The nodes represent noun categories (e.g., dog, telephone, spoon) that are produced between 16 and 30 months of age (or between 1.5 and 2.5 years); these categories, then, are the formative base of the human category system. The links are formed when nouns share features derived from adult feature generation norms. Thus, the resulting networks are organized according to feature correlations.

There are two potential criticisms of this approach. First, a limitation of using feature norms (well-recognized in the feature generation literature, Cree & McRae, 2003; McRae, Cree, Seidenberg, & McNorgan, 2005) is that the features provided by adults do not usually include crucial but not easily labeled properties (e.g., "cow shaped") nor properties so essential that they are apparently assumed and not mentioned (e.g., "breathes"). In the present case, this limitation implies that our results should be taken as a conservative estimate of what children could infer from the available feature information, as the missing features seem likely to license more robust categorical inferences. Second, it is possible that children do not have access to all the features listed by adults. To address this, we consider only a subset of adult generated features, using only perceptual and conceptual features that characterize everyday experiences with these things: dogs have fur, four legs, and bark; airplanes have wings, fly, and are made out of metal; apples are sweet, have seeds, and grow on trees, etc.

For the present analyses, the features were taken from the feature norms reported by McRae et al. (2005). Cree and McRae (2003) classified these features into mutually exclusive kinds. Two of those kinds, in their classification system, were called (1) perceptual features (e.g., "has 4 legs", "has a tail"), which refer to stable and perceivable properties of a thing and (2) functional features (e.g., "used for racing", "used for transportation"), which refer to how a thing may be used or its role in an event. These two kinds of features were defined independently of the goals of this study, but also overlap with the perceptual-conceptual distinction in the developmental literature. Accordingly, we use them – as given by Cree and McRae (2003) – to examine whether so-called perceptual and conceptual features differ in their contributions to children's categories, or perhaps play similar roles (see Yoshida & Smith, 2003).

This is primarily a descriptive study that addresses four specific questions: (1) Do features provide sufficient structure to infer common adult taxonomic categorizations among the nouns children know at 30 months of age? (2) If so, what are the available categories? (3) How robust are these categories to more or less stringent criteria for feature correlations? (4) Do perceptual and conceptual features differ in the structure they provide – are some feature types more robust, more discriminating, or more redundant?

2. Methods

2.1. Noun categories

The nouns were selected from the MacArthur-Bates Communicative Developmental Inventory (Fenson et al., 1994), Toddler version. This inventory contains the words that were in at least 50% of children's productive vocabulary by 30 months in a large normative study. Feature norm data are available (from McRae et al., 2005) for 130 nouns (a subset of the 312 nouns on the MCDI). These 130 nouns over-represent (with respect to the inventory as a whole) animals (33 nouns, 25% of the subset versus 15% of whole inventory) and under-represent food (17 nouns, 13% of the subset versus 23% of the whole inventory). Nonetheless, the sample includes a broad array of nouns across several different superordinate categories. The complete list of nouns is given in Appendix.

2.2. Features

The features were taken from the feature norms reported by McRae et al. (2005). That study collected features for 541 nouns from 725 adults with 30 adults providing features for each noun. The participants were given a noun and 14 blank spaces to fill with features. They were prompted to provide physical properties (how it looks, smells, sounds, etc.), functional properties or uses, internal properties, and other pertinent facts. We use the brain region coding presented in Cree and McRae (2003) to classify the features. This classified the features into 4 perceptual feature sets representing the 5 senses (e.g., "is yellow", "is soft"), functional (e.g., "eaten by monkeys", "eaten by peeling"), encyclopedic (e.g., "grows in tropical climates") and taxonomic (e.g., "a fruit"). We used only features coded by Cree and McRae as perceptual and functional (conceptual in our usage) for three reasons. First, these are the two kinds of features about which developmental theories have been concerned. Second, these kinds of features are likely to be in the everyday experiences of young children. Third, superordinate names (e.g., "a fruit") – the likely real-world correlate of taxonomic features in Cree and McRae's classification (2003) – are not typically known by children younger than three years of age. Also, our principle focus concerns the ability of feature correlations among perceptual and conceptual features to form a representational basis for higher-order categories later in life.

2.3. The networks

Nodes represent nouns and edges represent features that are shared between nouns. To investigate how the quantity of shared information between nouns influences categorical structure, we define edges in terms of differing numbers of shared features. For example, when w (the feature threshold to define an edge) is 1, nouns are connected by an edge if they share at least one feature, and when w is 2, nouns are connected by an edge if they share at least 2 features. These different criteria for defining edges (the connectedness of any two nouns) yield a series of networks, which correspond to different requirements for shared numbers of features, with larger w meaning more information is required for connectedness.

In total, the analyses are based on 130 nouns. The total number (tokens) of features associated with these nouns is 1394, with 991 perceptual tokens and 403 functional tokens. The number of features per noun ranged between 6 and 17 ($M = 11.08$, SD = 2.4). The number of unique features (types) were 655: 385 perceptual and 270 conceptual. Because we only use shared features to detect categories (as proposed by Rosch et al., 1976), many of these features do not contribute to the network structure as they occurred with just one noun, and they are not included in the subsequent analyses. One hundred and ninety-nine unique features were shared by at least 2 nouns. These consisted of 57 conceptual features and 142 perceptual features, consisting of 1 smell feature, 3 sound features, 13 tactile features, 4 taste features, 13 visual-color features, 97 visual-parts features, and 11 visual-motion features.

2.4. Small-world analyses

Watts and Strogatz (1998) measured small-world structure by comparing the average clustering coefficient for the network being analyzed with that expected for a randomly connected control network (a network that has the same number of nodes and edges, but with randomly assigned connections). The clustering coefficient for each node is calculated by determining the proportion of a node's closest neighbors (nodes connected by an edge) that are also connected to each other by an edge. For example, Fig. 1 demonstrates the clustering coefficient calculated for three separate nodes. The clustering coefficient of a node, e.g. $c(a)$, is calculated by determining how many connections exist between nearest neighbors of that node (node a). The number of possible connections that can exist between neighbors is determined by the node's degree: $\tau(a) = \frac{d(a)^2 - d(a)}{2}$. The clustering coefficient is then the fraction of observed connections, $\lambda(a)$, among those possible: $c(a) = \frac{\lambda(a)}{\tau(a)}$.

To get the clustering coefficient for the network as a whole, the clustering coefficient is averaged for all nodes. When a network has a high average clustering coefficient relative to the appropriate random control network, it indicates the existence of subnetworks, or clusters of *kinds of categories*. We use this measure to ask how well features organize early-learned categories into clusters of higher-order categories and more specifically, the role of perceptual and conceptual features in that organization.

2.5. Cluster analyses

To identify categories in a principled way, we sought a method that does not force items into categories (in contrast to hierarchical clustering algorithms). That is, TELEPHONE and BOOK may properly not belong to any "superordinate category" in a young child's semantic knowledge, because they do not share enough features with other nouns (and may not therefore support generalizations to or from other artifacts). We also wanted to avoid forcing objects into only one category, because items
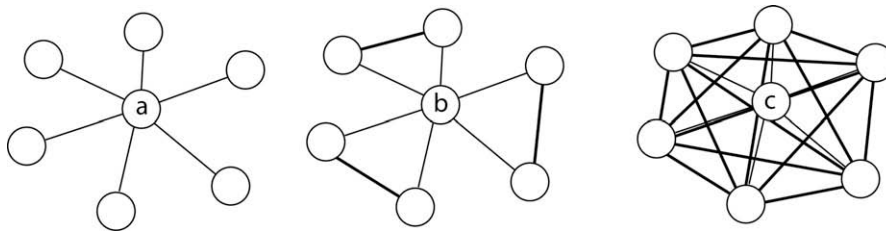
Fig. 1. Graphs and clustering coefficients: $c(a) = 0/15 = 0$, $c(b) = 3/15 = 0.2$, $c(c) = 15/15 = 1.0$.
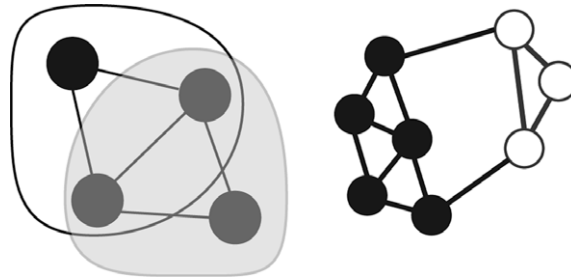

Fig. 2. Clique percolation method. On the left, two 3-cliques (circled) are 3-clique connected because they share two neighbors. On the right: a network showing two clique percolation clusters (black and white), which are composed of 3-clique connected subgroups.

may well belong to more than one category. For example, AIRPLANE may belong in a category with flying things, but it may also share features with other vehicles. Given these goals, we used the clique percolation method introduced by Palla, Derenyi, Farkas, and Vicsek (2005).

The clique percolation method identifies groups of nodes that are well connected with one another. It does this by identifying the presence of cliques, which are sets of nodes that are all connected with one another (maximal complete subgraphs). A k-clique represents a set of k nodes where all k nodes are connected to one another. Two k-cliques are adjacent if they share $k - 1$ vertices (see Fig. 2). Two k-cliques are k-clique-connected if they are connected by a sequence of adjacent k-cliques. A k-clique percolation cluster is the union of all k-cliques that are k-clique-connected to one another. In the present case, the clique percolation method identifies nouns that share sufficient feature correlations (sensu Rosch et al., 1976; Rogers & McClelland, 2004), or that are sufficiently connected through other nodes, to be considered clusters.

For a given value of k, the clique percolation method identifies all k-clique percolation clusters. Fig. 2 illustrates the method showing, on the left, how two sets of 3 nodes ($k = 3$) are 3-clique connected because they share 2 ($k - 1$) edges and on the right showing two clique percolation clusters (composed of 3-clique connected subgroups).

The clique percolation method also provides a principled approach to identifying the cut-off threshold that yields the most structural information (see Palla, Barabasi, & Vicsek, 2007). This is accomplished by increasing the value of k for each cut-off threshold, w, until the second largest component is larger than half the size of the largest component. For low values of k, most nodes tend to be connected in one large cluster. However, as k is increased, the

percolation clusters separate as the method focuses in on narrow regions of high connectivity. After adjusting k upwards for each cut-off threshold, we then identify the corresponding w and k that have the largest number of percolation clusters, and therefore the most putatively identifiable categories.

3. Results

The analyses take the following approach: First, we ask if the full network of features provides sufficient information to infer the higher-order categorical structure among the words that children are likely to know at 30 months of age, and if so, what are the higher-order categories likely to represent. Second, we examine the relative contribution of conceptual and perceptual features, again asking how structurally informative these features are, and what categories they give access to.

3.1. The full network

3.1.1. Network statistics and small-world analyses

Fig. 3 shows a series of noun–feature networks, with nouns connected if they share at least 1, 2, 3, or 4 features ($w = 1$, 2, 3, or 4, respectively).[1] When $w = 1$, there is one densely connected network. When $w = 2$, 3 and 4, subgraphs emerge and considerable structure is apparent. Visual inspection reveals that nouns that refer to animals tend to be connected to each other, nouns that refer to foods are connected to each other, and so-forth. The clusters of nouns

---

[1] We limit our analysis to w between 1 and 4, because above $w = 4$ more than half the nodes are isolates in the full network.
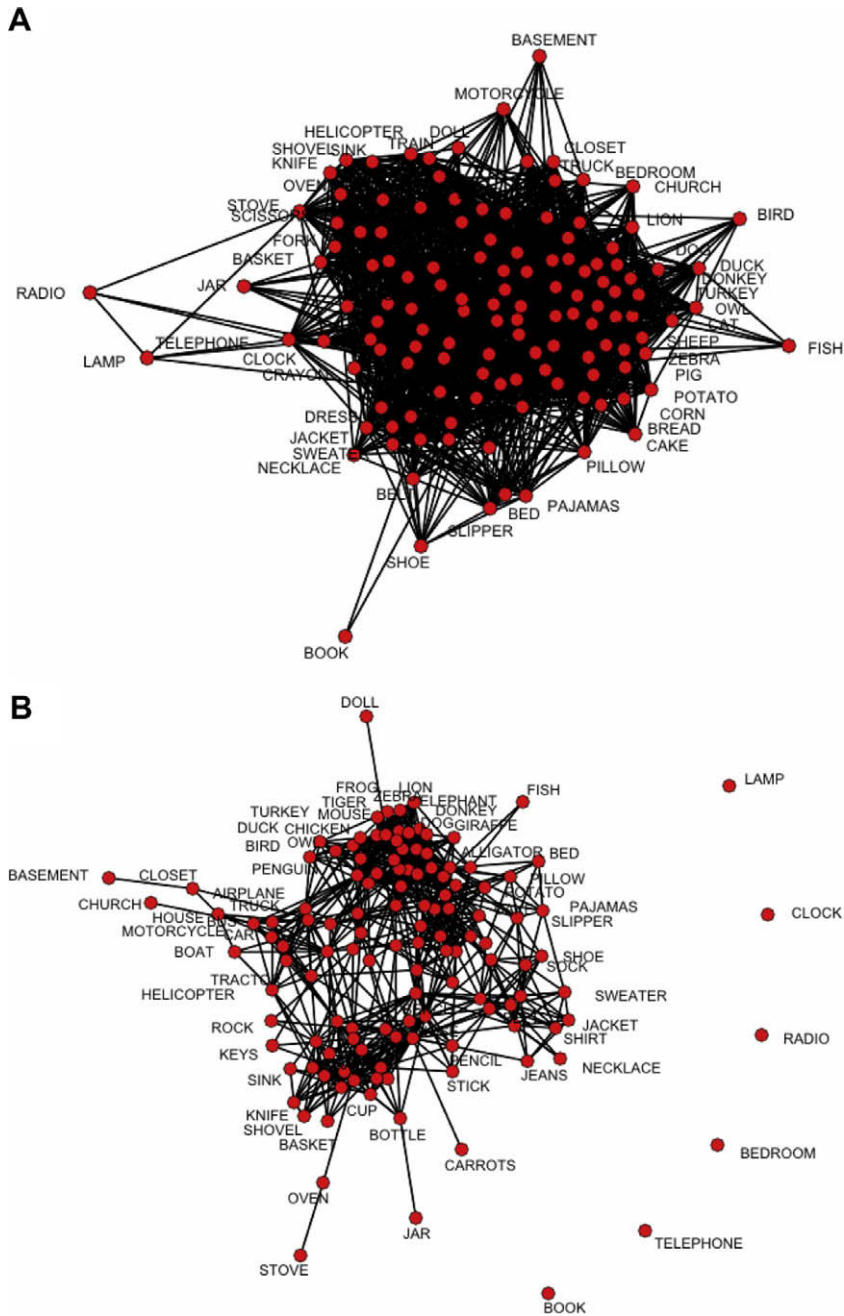
Fig. 3. A series of networks of the 130 nouns with each noun connected to another noun if they shared one feature (A), at least 2 features (B), at least 3 features (C), and at least 4 features (D).

that share the most correlated features (apparent in the $w = 4$ network) are animals, vehicles, foods, clothes, and household objects. Thus, features alone can represent categorical information, but increasing the threshold for the number of features required to produce an edge leads to more meaningful category subdivisions.

To formalize the existence of subnetworks formed by shared features, we calculated the average clustering coefficient across all nodes in each network (see Fig. 1). This is then compared to the mean clustering coefficient for 500

randomly connected networks with the same density (i.e., number of nodes and edges). The clustering coefficient (C), average shortest path length (L), and related graph statistics are reported in Table 1.

Table 1 reveals that the noun–feature network of the nouns normatively known at 30 months has all the properties of a small-world network ($L \approx L_{random}, C \gg C_{random}$). This property is also robust to increasing values of the cut-off threshold, w. As w increases from 1 to 4, the clustering coefficient increases from 0.55 to 0.6, while the average

Fig. 3 (*continued*)

clustering coefficient of the 500 random networks of the same density goes from 0.29 to 0.02.[2] The presence of

small-world structure in the noun–feature network is consistent with the structure observed for other semantic and real-world networks (Steyvers & Tenenbaum, 2005; Watts & Strogatz, 1998). With respect to early concept development in children, the small-world structure provides a basis for superordinate categorical structure; it has the following properties: (1) some items are located in robust clusters (i.e., even when the number of shared features required for an edge is high), (2) some items are not found in categorical clusters (these are the isolates), and (3) because of the nat-

---

[2] The difference in the distributions of the ratios of observed to random clustering coefficients is significantly different between thresholds (data not shown). However, at present there is no quantitative criteria for stating that one network is more or less of a small-world than another. Our interpretation of the data is that the categorical structure is robust to changes in feature threshold. For example, for the full network, significant structure is observed out to $w = 6$, which has a clustering coefficient of 0.27, and categories of food, vehicles, clothes, animals, and furniture are still visible (data not shown).

Table 1
Statistics for the full-network when at least 1, 2, 3 or 4 shared features is required to connect any two noun categories. Columns represent the following: (1) clustering coefficient; (2) average shortest length to connect every possible pair of nodes *within* a component (the within component criteria allows for the nonmonotonic progression in lengths); (3) mean (with standard deviations in parentheses) of the clustering coefficient computed for the random networks; (4) mean (with standard deviations in parentheses) of the average path length computed for the random networks; (5) density – observed number of edges divided by possible number of edges; (6) clusters represent the number of *unconnected* components that contain at least 2 nodes; (7) isolates, the number of nodes that are not connected to any other node. *Indicates a significant difference ($p < 0.001$) for the clustering coefficient, from the random population, using a one-sample t-test.

w	Clustering coefficient	Average length	Random clustering coefficient	Random average length	Density	Clusters	Isolates
1	0.55*	1.76	0.29 (0.01)	1.71 (0.01)	0.29	1	0
2	0.54*	2.57	0.10 (0.01)	2.15 (0.03)	0.10	1	6
3	0.58*	4.96	0.04 (0.01)	3.12 (0.09)	0.04	4	20
4	0.60*	3.22	0.02 (0.01)	4.89 (0.38)	0.02	10	47

Table 2
Clique percolation clusters from the full network. Superscripts refer to category designations provided in the MCDI (Fenson et al., 1994). The second occurrence of a taxonomic label is indicated by an acronym taken from the first letters of the label.

BREAD[FOOD AND DRINK] CHEESE[FAD] LAMB[ANIMAL]
BALLOON[TOY] BENCH[FURNITURE AND ROOM] BOWL[SMALL HOUSEHOLD ITEM] BROOM[SHI] BRUSH[SHI] BUCKET[SHI] COMB[SHI] CUP[SH] DISH[SHI] FORK[SHI] HOSE[OUTSIDE THING] PEN[T] PLATE[SHI] SLED[VEHICLE] SPOON[SHI] TRAY[SHI]
DOOR[FAR] HAMMER[SHI] SHOVEL[OT] SPOON[SHI]
FORK[SHI] KNIFE[SHI] SCISSORS[SHI]
BEAR[A] CAT[A] COUCH[FAR] COW[A] DEER[A] DOG[A] DONKEY[A] ELEPHANT[A] FROG[A] HORSE[A] LAMB[A] LION[A] MOOSE[A] MOUSE[A] PIG[A] PONY[A] SHEEP[A] SQUIRREL[A] TIGER[A] TURTLE[A] ZEBRA[A]
BOOTS[CLOTHING] COAT[C] DRESS[C] JACKET[C] JEANS[C] PANTS[C] SCARF[C] SHIRT[C] SHOES[C] SLIPPERS[C] SOCKS[C] SWEATER[C]
APPLE[FAD] BANANA[FAD] BEANS[FAD] CAKE[FAD] CHEESE[FAD] GRAPE[FAD] ORANGE[FAD, DESCRIPTIVE WORD] PEAS[FAD] PICKLE[FAD] RAISIN[FAD] STRAWBERRY[FAD] TUNA[FAD]
BEAR[A] BIRD[A] CHICKEN[A, FAD] DUCK[A] GOOSE[A] OWL[A] PENGUIN[A] ROOSTER[A] TURKEY[A]
BED[FAR] CHAIR[FAR] COUCH[FAR] PILLOW[SHI] SLIPPER[C] SOFA[FAR]
AIRPLANE[V] BICYCLE[V] BUS[V] CAR[V] HORSE[A] MOTORCYCLE[V] TRACTOR[V] TRAIN[V] TRICYCLE[V] TRUCK[V]

ure of small-worlds, some items provide cross-overs between clusters, which keeps the average path length low, even when all items are connected ($w = 1$).

The observation of small-world structure offers testable predictions. For example, in these networks, three or fewer features connect relatively many noun categories but only a few categories are connected by at least 4 features. If connectedness in these networks is predictive of psychological similarity then these more highly interconnected subgraphs (in the $w = 4$ network) should be expected to better support generalizations from one basic-level category to another, compared with subgraphs formed under lower thresholds (e.g., when $w = 3$, 2, or 1). Similarly, basic-level categories that are the first to become isolated as w increases (categories such as BOOK and TELEPHONE) may be the least likely to support such generalizations.

3.1.2. Cluster analyses

The above analyses based on the clustering coefficient indicates the existence of subnetworks of local structure. What are these clusters and how coherent are they? To identify these, we used the clique percolation method described above (Palla et al., 2005). For the noun–feature network, the k and w values that yield the most clusters are 3 and 3, respectively. This yields a conservative estimate for category membership, because only nouns with enough local information to be included in a clique of size $k = 3$ will be included in the output. Nouns lacking this connectedness are not assigned to any cluster. The 10

clusters identified for these values of k and w are listed in Table 2.[3]

These clusters represent potential category structure and are generally consistent with our adult expectations, at least in terms of what they include. We provide as superscripts the category designations provided in the MCDI, which we consider to be reasonable estimates of how adults would organize these words.[4] Comparing these taxonomic memberships with the percolation clusters finds significant parallels. Categories that are perfectly consistent with adult taxonomic categories – in terms of what they include – are Food and Drink, Vehicles, and Clothing. The clique percolation method using feature overlap identifies these categories with no errors of inclusion; there is nothing present that doesn't belong. It is also interesting that the feature clusters pick up ad hoc categories (Barsalou, 1983) such as a category of ITEMS FOR CUTTING, a SOFT-WHITE THINGS category, and a category of THINGS TO REST AND RELAX. However, in some cases, category members lie outside our intuitive taxonomic assignments. For example, COUCH is in a category with animals, because it "has four legs", "is large",

---

[3] While the clique percolation method will identify clusters in any network, the absolute values of k and w are relative to the edge information provided; in the present case they are relative to the information provided in the adult feature norms.

[4] We chose the MCDI designations over those provided by Cree and McRae (2003), because they were identified independently of the features produced in the feature generation norms. The MCDI categorizations also include fewer singletons. However, using the Cree and McRae categories does not alter our interpretation or the conclusions we draw.

and "is soft". COUCH is also an item that is found in more than one category, as are five other items: LAMB, FORK, SPOON, BEAR, and HORSE. Most of these are arguably correct (except BEAR in the birds category), but as we note in the following section, overgeneralization and errors of inclusion are but one end of a trade-off between generalization and specificity.

In sum, the results from the full network demonstrate the following: First, readily available features among nouns that children know at 30 months provide sufficient information to structure these nouns into superordinate categories, without the use of taxonomic labels. Second, the structure (shown in Table 1) provided by feature information is robust to perturbations in the number of features required to form a categorical relationship. This indicates that the necessary small-world structure required to produce meaningful categories is largely redundant and therefore robust to random feature omissions. And third, the categories that do arise out of feature overlap are to a large extent exactly those categories adults consider reasonable when categorizing these nouns a priori. In the following section we examine each of the two main feature types to determine how each provides structure in the full network.

3.2. Perceptual and conceptual feature networks

3.2.1. Network statistics and small-world analyses

To represent the kinds of semantic structural information children would have if they used only conceptual or perceptual relatedness to link categories, we composed separate networks of only perceptual or conceptual features. Figs. 4 and 5 present the conceptual and perceptual networks at the thresholds found to reveal the most structure via the clique percolation method. Table 3 presents statistics for the series of perceptual and conceptual networks for $w = 1–4$.

As is apparent from Figs. 4 and 5 and Table 3, the perceptual network is far denser than the conceptual network. On average, a node in the perceptual network at $w = 1$ is connected to 27% of the other nodes; the average node at the same cut-off threshold in the conceptual network is only connected to 5% of the other nodes. This would suggest that conceptual information is more discriminating than perceptual information among these early-learned nouns.

The discriminatory role of conceptual information is also evident in the number of nouns to which the features link. The most common conceptual features (in terms of the number of nouns with which they are associated) are: "is edible" (20), "used for transportation" (11), "worn for warmth" (8), "hunted by people" (6), "used by children" (6), and "used for holding things" (6). The most common perceptual features are: "made of metal" (24), "different colors" (22), "has four legs" (22), "is large" (21), and "is small" (21). Note that perceptual features divide the nouns in to two very large categories: small and large objects (metal) and small and large living entities (4 legs) where functional features divide the world into more categories. Note also that the most common perceptual features are more promiscuous (appear with more nouns) than the most common conceptual features. Across all nouns, con-
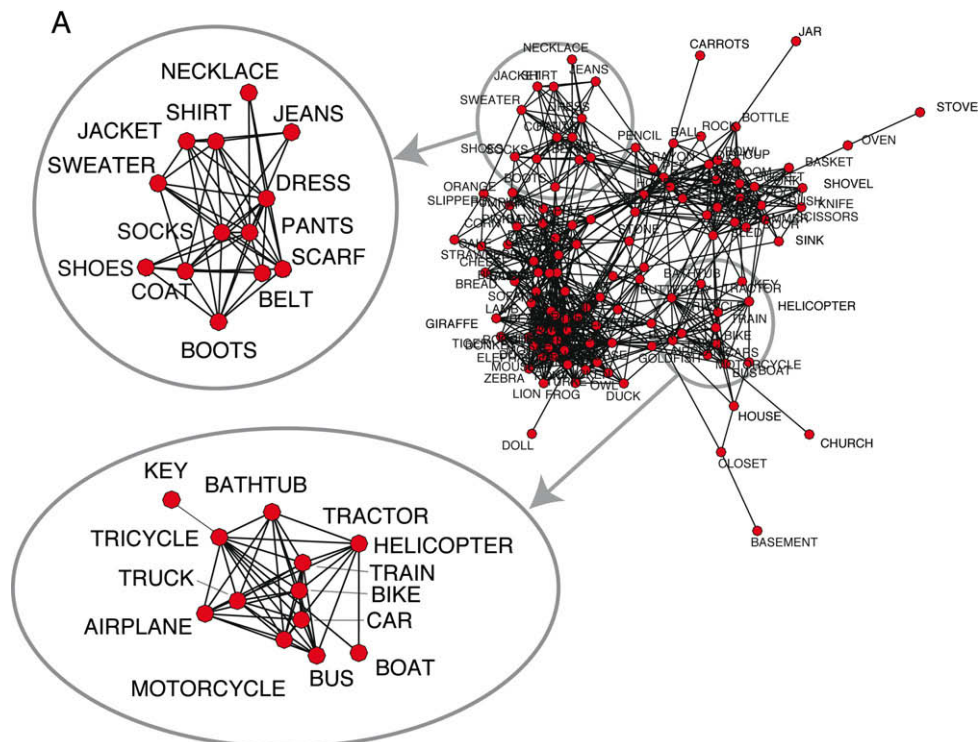

Fig. 4. The perceptual network with $w = 2$. Grey circles indicate areas of the larger network that are enlarged for clarity.
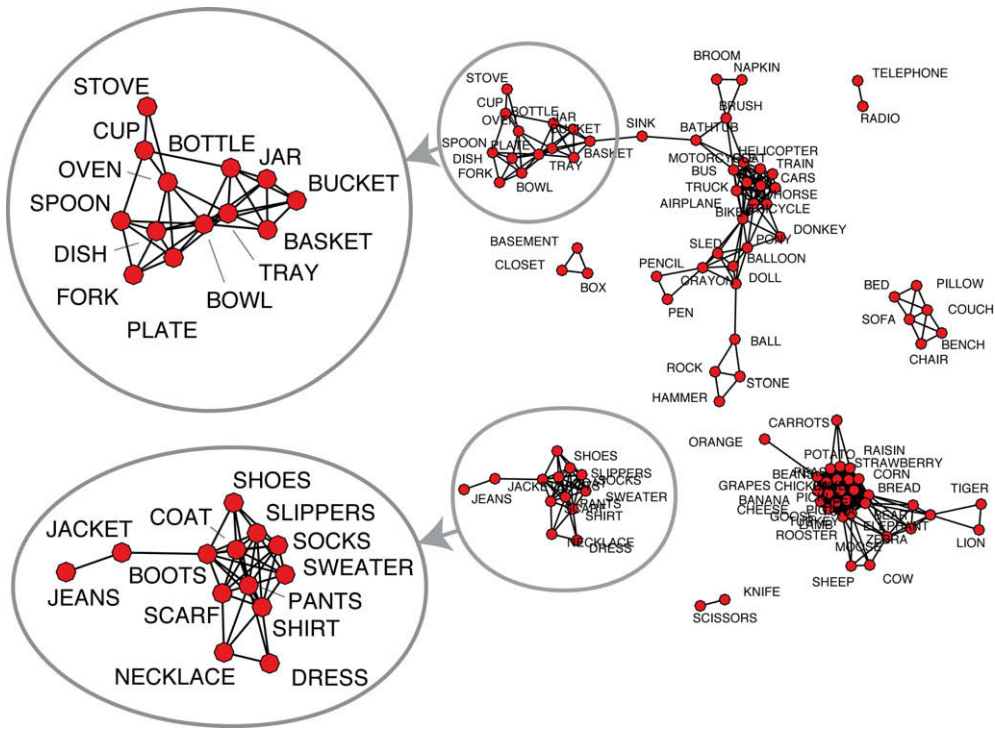
Fig. 5. The conceptual network with $w = 1$. Grey circles indicate areas of the larger network that are enlarged for clarity.

Table 3
Statistics for the perceptual-feature network (up) and the conceptual-feature network (down) when at least 1, 2, 3 or 4 shared features is required to connect any two noun categories: Columns are as in Table 1. *Indicates a significant difference ($p < 0.001$) for the clustering coefficient from the random population, using a one-sample t-test.

w	Clustering coefficient	Average length	Random clustering coefficient	Random average length	Density	Clusters	Isolates
Perceptual networks							
1	0.54*	1.76	0.27 (0.005)	1.73 (0.005)	0.27	1	0
2	0.54*	2.69	0.08 (0.01)	2.33 (0.03)	0.08	2	10
3	0.62*	5.02	0.03 (0.01)	3.66 (0.16)	0.03	8	31
4	0.62*	2.55	0.01 (0.02)	6.17 (1.55)	0.01	11	69
Conceptual networks							
1	0.88*	3.08	0.05 (0.01)	2.80 (0.06)	0.05	7	33
2	0.59*	1.59	0.01 (0.02)	6.31 (1.67)	0.01	14	81
3	0.38	5.02	0.43 (0.49)	1.10 (0.12)	0.00	7	113
4	1.00	1.00	0.94 (0.21)	1.00 (0.06)	0.00	2	126

ceptual features share on average 1.54 (SD = 1.64) nouns and perceptual features share 2.58 (SD = 3.67) nouns. The results of a Wilcoxon rank sum test show these differences are significant (W(43245), $p < 0.001$); per feature, conceptual features are associated with fewer nouns than perceptual features.

The discriminating nature of conceptual features has the further consequence that the number of isolates is much higher for the conceptual network than for the perceptual network. At a cut-off threshold of $w = 2$, more than half of the nodes in the conceptual network are unconnected to any other node. At the same cut-off threshold for the perceptual network, only 10 nodes are isolates. The greater increase in isolates for the conceptual network arises for two primary reasons. One is that most animals

have no functional features. The other is that most objects are used for one main function only. This has the consequence that shared perceptual features tend to be more redundant than conceptual relationships – perceptual features can be removed with less radical structural alteration of the network. Indeed, edge relationships in the conceptual network are predominantly based on a single shared feature.

When using all perceptual and conceptual features, both networks have small-world structure. With w ranging from 1 to 4, the conceptual network clustering coefficients range from 0.88 to 1. For the same w range, the perceptual network clustering coefficients range from 0.54 to 0.62. Using the clustering coefficient as a measure of local structure, one conceptual feature is apparently as good as or

better than 4 perceptual features in creating that structure. However, at $w = 2$, the number of isolated nouns in the conceptual network is 81, but only 10 for the perceptual network. Thus, while conceptual networks appear to be more discriminating, they are also more sensitive to the presence or absence of any given feature. Conceptual features appear to trade-off robustness for precision, while perceptual features are more robust but less precise.

The argument that conceptual features are more discriminating and thus potentially more effective at isolating categories is further evidenced by the fact that the difference between the observed clustering coefficients and that for a random network of similar density is higher for the conceptual network than for the perceptual network. This is consistent with what we can visually observe in Figs. 4 and 5: the conceptual network has more local structure than the perceptual network. However, even the slightest increase in the cut-off threshold reduces the conceptual network to a large number of isolates. Meanwhile, the perceptual network maintains small-world structure and involves the majority of the nodes in this structure even if the requirement for noun-pair relatedness is three or more perceptual features.

The above analyses reveal that perceptual features (as provided by adults) are more robust to changes in the underlying threshold. However, this may be due to there simply being more perceptual features in the feature generation norms. To control for this, we created 200 perceptual subnetworks, where for each subnetwork we randomly selected as many perceptual features as there are conceptual features. Table 4 presents the statistics for these 200 perceptual subnetworks and shows where they are significantly different from the matched conceptual networks at each threshold. The results clearly indicate that feature-for-feature, perceptual features do far less work at organizing categorical information. There are more isolates and fewer clusters for the perceptual subnetworks.

While the clustering coefficient appears higher at $w = 3$, this does not control for the number of nodes still connected in the network. To control for this, we computed the normalized clustering coefficient, which is the clustering coefficient multiplied by the fraction of nodes that are not isolates. Fig. 6 presents the normalized clustering coefficient for each threshold value for each of the network representations. It clearly shows that perceptual features, when matched to the number of conceptual features, are significantly less effective at clustering nouns than conceptual features. Note also, however, that the normalized clustering coefficient for the full perceptual network is similar
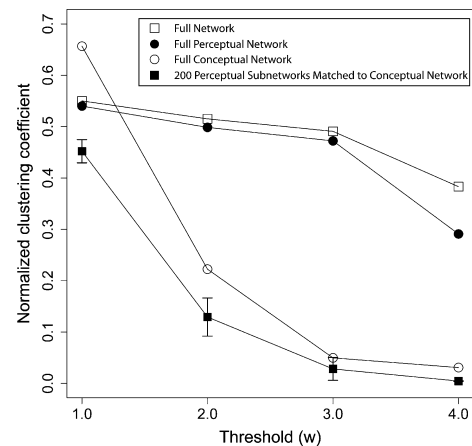

Fig. 6. The normalized clustering coefficient for each threshold presented for the various networks. The error bars on the 200 perceptual subnetworks represent the 95% confidence interval.

to that of the full network, and appears to drive most of the categorical structure in the full network. Thus, perceptual information may provide the lion's share of information relevant to category inference, but this appears to be due to their abundance, not because individual perceptual features are more informative.

Taken together, these results support the idea as proposed by many (Keil, 1989; Mandler, 1992; Carey, 1985; Gelman, 1990) but doubted by others (Ahn & Luhman, 2005; Smith, 2005) that perceptual and conceptual features contribute differently to category organization. Further, there is a clear trade-off here. Perceptual information, because of its abundance, is more redundant and can provide more robust information about category inclusion, but this information is not as discriminating as conceptual information. A single conceptual relation is sufficient to define all category members that are, for example, "used for transportation." No single perceptual feature contains that information.

3.2.2. Cluster analyses

Using the clique percolation method, the conceptual network provides the most number of clusters (11) when $k = 3$ and $w = 1$; for the perceptual network, the most clusters (9) are separated out when $k = 5$ and $w = 2$. This is consistent with the graph-theoretic data in Table 3 showing that the conceptual network has fewer isolates and greater local structure at its lowest cut-off threshold, while the

Table 4
Statistics for 200 perceptual subnetworks, composed of randomly selected perceptual features. The number of perceptual features was chosen to match the number of conceptual features produced in the feature norms. Columns are as in Table 1. *Indicates a significant difference ($p < 0.001$) from the matched conceptual networks, using a one-sample t-test.

w	Clustering coefficient	Average length	Density	Clusters	Isolates
200 perceptual subnetworks					
1	0.52*	2.79*	.062*	1.9*	17.3*
2	0.42*	2.14*	.00*	10.12*	90*
3	0.59*	1.15*	.00	2.6*	123.8*
4	0.98	1.03	.00	0.27*	129.4*

perceptual network loses only a few nodes to isolates but gains substantial local structure – compared with a random network of the same density – by increasing w to 2.

A close look at Tables 5 and 6 and the different kinds of clusters present in the two networks reveals some interesting comparisons. First, there is a difference in cluster size between the two groups. Clusters in the conceptual network are generally smaller ($M = 8.45$, Median = 4) than those in the perceptual network ($M = 13$, Median = 12). The conceptual categories also appear to be more conservative – there are fewer odd members in any category. Using a liberal inclusion method – where an object is included in a superordinate category if we can imagine any argument in favor of its inclusion – we count 3 odd objects among the conceptual clusters and 9 among the perceptual clusters (e.g., CRAYON, DOLL, and BRUSH, are in the dominantly vehicle category among the conceptual clusters, while HOSE and PEN are in the dominantly fruit category among the perceptual clusters). Using the MCDI category labels, the only unmixed category among the perceptual features is vehicles, while conceptual features provide four unmixed categories, consisting of small household items (for cleaning), toys (for drawing), animals, and clothes. Finally, we note the number of items in more than one category differs between the two feature types: The perceptual categories have 11 nouns that are in more than one category, whereas there are only 4 duplicate nouns in the conceptual categories. Compared with the 6 duplicate nouns in the full network, it is again clear that perceptual features are more inclusive in determining category memberships, compared with conceptual features.

We warn against blaming these category inclusion errors on the clique percolation algorithm. It can only use the information it is provided with, and it does quite well when provided with all features, and in other paradigms (see Palla et al., 2005). Also, though an individual category inclusion error may be argued one way or the other, we feel the weight of the evidence provided above shows that perceptual features do overgeneralize category boundaries at the risk of inclusion errors, whereas conceptual features appear to do just the opposite.

Finally, compared with the full network, both features types produce categories more representative of *ad hoc* categories. For example, to our best approximation, two of the perceptual clusters represent LONG THIN THINGS and THINGS THAT CAN FLY, plus there is a large category of ARTIFACTS held together because they are MADE_OF common materials like METAL and PLASTIC. Similarly for conceptual clusters, we find PLACES TO STORE THINGS, ITEMS FOR CLEANING, ITEMS FOR DRAWING, AND ITEMS FOR THROWING OR HITTING.

Table 7 provides a summary of the observed differences between the conceptual and perceptual networks. The conceptual networks typically involve fewer features, they are less dense, some categories are left out, they are less likely to put items in more than one category, and they are not robust to the omission of features. However, individual categories are well discriminated and are more likely to in-

Table 5

Clique Percolation Clusters from the Perceptual Network. Superscripts refer to category designations provided in the MCDI (Fenson et al., 1994). Acronyms are as in Table 2.

BALL[T] BALLOON[T] BASKET[SHI] BENCH[FAR] BOWL[SHI] BROOM[SHI] BRUSH[SHI] BUCKET[SHI] COMB[SHI] CUP[SH] DISH[SHI] DOOR[FAR] FORK[SHI] HAMMER[SHI] HOSE[OT] KNIFE[SHI] PEN[T] PLATE[SHI] SCISSORS[SHI] SHOVEL[OT] SLED[V] SPOON[SHI] TRAY[SHI]
BED[FAR] COUCH[FAR] LAMB[A] PAJAMAS[C] PILLOW[SHI] SLIPPERS[C] SOFA[FAR]
BELT[C] BENCH[FAR] COAT[C] DRESS[C] JACKET[C] JEANS[C] PANTS[C] PEN[T] SCARF[C] SHIRT[C] SOCKS[C] SWEATER[C]
APPLE[FAD] BANANA[FAD] BEANS[FAD] GRAPE[FAD] HOSE[OT] ORANGE[FAD,DW] PEAS[FAD] PEN[T] PICKLE[FAD] PUMPKIN[FAD] RAISIN[FAD] STRAWBERRY[FAD]
BREAD[FAD] CHEESE[FAD] LAMB[A] NAPKIN[SHI] PILLOW[SHI]
BELT[C] BENCH[FAR] BROOM[SHI] HOSE[OT] PEN[T] PENCIL[T] STICK[OT]
BUS[V] CAR[V] MOTORCYCLE[V] TRACTOR[V] TRAIN[V] TRUCK[V]
AIRPLANE[V] BIRD[A] BUTTERFLY[A] CHICKEN[A, FAD] COW[A] DUCK[A] GOOSE[A] OWL[A] PENGUIN[A] ROOSTER[A] TURKEY[A]
BEAR[A] BOX[SHI] CAT[A] CHAIR[FAR] CHICKEN[A, FAD] COUCH[FAR] COW[A] DEER[A] DOG[A] DONKEY[A] ELEPHANT[A] FROG[A] GIRAFFE[A] HORSE[A] LAMB[A] LION[A] MOOSE[A] MOUSE[A] PENGUIN[A] PIG[A] PONY[A] RAISIN[FAD] SHEEP[A] SOFA[FAR] SQUIRREL[A] TABLE[FAR] TIGER[A] TURKEY[A] TURTLE[A] ZEBRA[A]

Table 6

Clique Percolation Clusters from the Conceptual Network. Superscripts refer to category designations provided in the MCDI (Fenson et al., 1994). Acronyms are as in Table 2.

BASEMENT[FAR] BOX[SHI] CLOSET[FAR]
BROOM[SHI] BRUSH[SHI] NAPKIN[SHI]
CRAYON[T] PEN[T] PENCIL[T]
CUP[SH] OVEN[FAR] STOVE[FAR]
ELEPHANT[A] LION[A] TIGER[A]
BALL[T] HAMMER[SHI] ROCK[OT] STONE[OT]
BANANA[FAD] BEANS[FAD] BEAR[A] BREAD[FAD] CAKE[FAD] CARROTS[FAD] CHEESE[FAD] CHICKEN[A,FAD] CORN[FAD] COW[A] DEER[A] ELEPHANT[A] GOOSE[A] GRAPE[FAD] LAMB[A] MOOSE[A] PEAS[FAD] PICKLE[FAD] PIG[A] POTATO[FAD] RAISIN[FAD] ROOSTER[A] SHEEP[A] STRAWBERRY[FAD] TUNA[FAD] TURKEY[A] ZEBRA[A]
BED[FAR] BENCH[FAR] CHAIR[FAR] COUCH[FAR] PILLOW[SHI] SOFA[FAR]
BOOTS[C] COAT[C] DRESS[C] NECKLACE[C] PANTS[C] SCARF[C] SHIRT[C] SHOES[C] SLIPPERS[C] SOCKS[C] SWEATER[C]
AIRPLANE[V] BALLOON[T] BATHTUB[FAR] BICYCLE[V] BOAT[V] BRUSH[SHI] BUS[V] CAR[V] CRAYON[T] DOLL[T] DONKEY[A] HELICOPTER[V] HORSE[A] MOTORCYCLE[V] PONY[A] SLED[V] TRAIN[V] TRICYCLE[V] TRUCK[V]
BASKET[SHI] BOTTLE[SHI] BOWL[SHI] BUCKET[SHI] DISH[SHI] FORK[SHI] JAR[SHI] OVEN[FAR] PLATE[SHI] SPOON[SHI] TRAY[SHI]

Table 7
Summary of the observed differences between conceptual and perceptual networks.

	Conceptual features	Perceptual features
Number	–	+
Density	–	+
Redundancy	–	+
Connected	–	+
Overgeneralizing	–	+
Items in Multiple Categories	–	+
Discriminating	+	–
Under generalizing	+	–
Information per feature	+	–

clude items that would be included in that category by adults. In contrast, the perceptual networks involve more features, are denser, hold their structure with less feature information, include most items in a category, are more likely to put items in more than one category, and are more likely to make errors of inclusion. In summary, conceptual categories tend to be smaller (underestimating category membership) and less sullied by near-members, whereas perceptual categories are larger and over-estimate category membership. These differences suggest that perceptual and conceptual features play distinct but possibly mutually supporting roles in category formation and use.

3.3. Correlations between conceptual and perceptual features

Although there are many differences between the conceptual and perceptual networks, they also – as is apparent in Table 5 and Table 6 – pick out overlapping, albeit not identical, higher-order clusters. These, then, are partially-redundant and correlated forms of category relatedness. We examined this overlap by considering a subset of 199 features (57 conceptual and 142 perceptual) that are present for at least 2 of the 130 nouns. We measured the degree to which these 199 features are associated with each other, defining association as the shared pattern of presence and absence across nouns. To compute this, we chose the Jaccard distance – also known as the asymmetric

binary distance – because it has the property that features present for the exact same nouns have a distance of 0 and features that are never present for the same noun have a distance of 1. The Jaccard distance for two features, a and b, takes the following form:

$$J(a,b) = 1 - \frac{n_{A \cap B}}{n_{A \cup B}}$$

where A is the set of all nouns sharing the feature a, B is the set of all nouns sharing the feature b, and n is the number of items in the set representing either the union or intersection of A and B. Classical multidimensional scaling was then used to transform the pairwise Jaccard distances into a two-dimensional set of coordinates so that the pattern of overlap between types of features could be visualized. In Fig. 7, the number 1 refers to perceptual features and the number 2 to conceptual features. We also list some of the specific features to make the apparent overlap more intuitive.

The figure shows a systematic relationship between perceptual and conceptual features. For any given conceptual feature, one is likely to find several perceptual features with roughly the same designations. So conceptual features and perceptual features share at least some of the work in the way they divide up the space. Moreover, because they are related with respect to the overlapping (if not exactly the same) higher-order categories, they provide two routes into higher-order categories, perhaps enabling children to bootstrap knowledge or inferences from one to another. Consistent with our previous analyses, perceptual features show more redundancy than the conceptual features – perceptual features more densely fill the MDS space in any given area, while conceptual features tend to be more evenly dispersed.

4. General discussion

The main contributions of the present analyses are as follows: (1) Perceptual and conceptual features commonly associated with nouns known by young children are sufficient to organize those nouns into small-world networks,
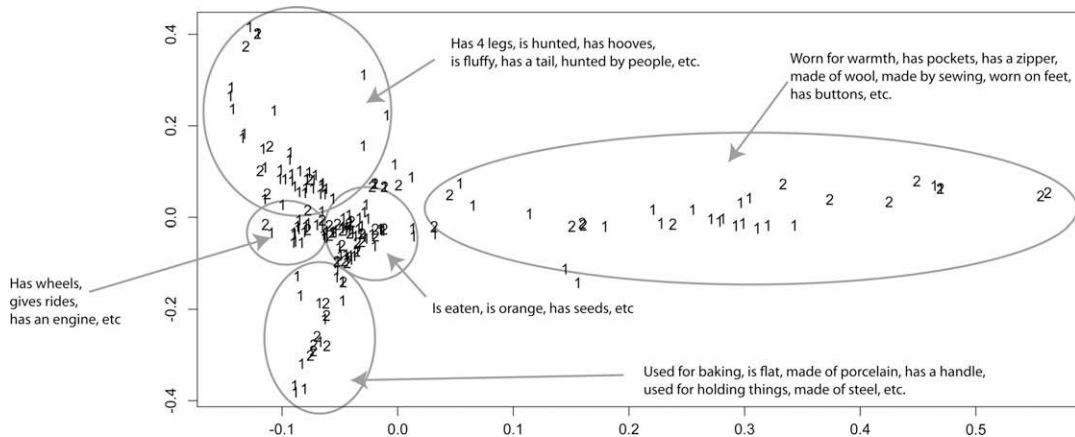

Fig. 7. The two-dimensional space derived from classical multidimensional scaling on the pairwise distance between features. The 1's are the perceptual features and the 2's are the conceptual features. There are several apparent clusters of features for each of these a subset of the feature labels are provided.

capable of representing higher-order categorizations. (2) These higher-order categorizations represent common superordinate categories as identified by adults. (3) These categorizations and the network structure underlying them are robust to minor changes in the criteria for category relations, but the degree of sensitivity to these changes is dependent on the kinds of feature involved. (4) Perceptual and conceptual features play different roles when structuring higher-order categories, with perceptual features being more abundant, more robust to random missing features, but less discriminating than conceptual features. In what follows, we discuss these contributions with respect to prior research in this area.

4.1. Higher-order categories from shared features

Following Rosch's (1973, 1975; Rosch et al., 1976) seminal papers and Smith and Medin's (1981) landmark book, the standard view of categories has been that while basic-level categories may be well-organized by overlapping and probabilistic features, superordinate, categories are not. Indeed, in the cognitive development literature, the existence of superordinate categories have been taken as *prima facie* evidence in favor of more abstract, more essentialist and theory-like representations of categories over representations in terms of mere feature distributions (Mandler, 1992; Gelman, 1990; Horton & Markman, 1980; Keil, 1994). The present results, however, show that shared features create clusters of categories rather like the traditional superordinate categories of food, clothing, animals, and so-forth. Things of the same general kind share correlated features. As Rosch et al. (1976) observed for basic-level categories, the world presents co-occurring properties that naturally group things into different kinds. This appears to be so for higher-order categories as well.

This conclusion fits the findings of McRae and colleagues, whose analyses of the feature distributions across adult categories also indicate superordinate groupings. Moreover, that work also shows that feature correlations predict adults' performance in a variety of category judgment tasks. The present results extend these findings by showing that higher-order categories may be derived from just the perceptual and functional features (without taxonomic or encyclopedic information) that are shared across a relatively small number of very early-learned basic categories. Thus, higher-order categories can be found in the feature correlations present at early stages of category development. The present results also fit with recent modeling efforts by Rogers and McClelland (2004) who also showed that feature correlations could generate superordinate categories. Their simulations of the incremental learning of these feature correlations also predicted observed developmental trends in a number of category judgment tasks. The present results go beyond these simulations (which were based on labeled links between categories and features that were generated by the theorists themselves) by showing that features normatively associated with the nouns children actually know early do have the requisite structure. In sum, although overlapping features may not be enough in and of themselves to explain all of human category organization, the present results suggest

that co-occurring features may be enough to start category knowledge off in the right direction.

4.2. Perceptual and conceptual features

Contemporary accounts of categories often distinguish between perceptual features and conceptual (relational/functional) features, with conceptual features assumed to be less probabilistic, more abstract, and the basis of higher-level distinctions (e.g., Sloutsky & Fisher, 2004; Gelman & Koenig, 2003; Keil & Batterman, 1984; see also Holyoak & Thagard, 1995; Hummel, 2000; Keil, 1989; Murphy & Medin, 1985). The observed differences between the networks built from perceptual versus conceptual features are consistent with and, indeed, provides a new form of support for this traditional view. A single shared conceptual feature yields well-organized and well-segregated superordinate groups, at least in terms of what the categories include. In contrast, the perceptual features yield messier approximations to these same categories, often overgeneralizing category memberships. Moreover, in terms of clustering per feature, conceptual features provide far more clustering information than perceptual features.

Many have hypothesized that category development proceeds from a more rough and probabilistic beginning to a more refined and essentialist mature structure (Gelman & Koenig, 2003; Keil & Batterman, 1984). The present finding that the more numerous and more redundant perceptual features are correlated with conceptual features across these early-learned categories could be used to support this view of perceptual features as the imperfect but critical starting point for superordinate category formation (Carey, 1985; Keil & Batterman, 1984; Sheya & Smith, 2006; Sloutsky & Fisher, 2004). Similar to Gentner's more general view that similarities in the surface properties of objects help learners discover relational structure (Kotovsky & Gentner, 1996), early attention to many overlapping perceptual properties, for example to redundant properties probabilistically characteristic of vehicles such as wheels, and doors, and seats, for example, could help children discover the more abstract and relational property of providing transportation.

We suspect that there is some truth to these ideas about category development. However, the larger framework may be wrong on two grounds. First, the assumption that the conceptual network is better than the perceptual network because its superordinate groupings are organized by a single conceptual feature may miss the cognitive importance of the full network structure. The full network has several properties characteristic of many real-world networks (including molecular, neural, semantic, and social networks: for a review, Barabasi, 2002; Csermely, 2006; Watts & Strogatz, 1998) that may be advantageous. For example, degeneracy is a property of many complex systems (for example, Edelman & Galli, 2001) in which a function can be accomplished in different ways by different components. In the full network, stable clusters of superordinate category organization emerge from different kinds of partially-redundant links. This is a form of degeneracy in that there is more than one way to form superordinate clusters and individual categories may be connected

to more than one of these clusters. In general, the value of degeneracy in a complex system is both increased stability (more than one way to the same outcome) and increased flexibility (variable paths). Weak links are also a common property of real-world networks; these are sparse long-range links between more densely connected subgroups and they appear to aid communication in the network and also enable the network –even one composed of well-articulated modules – to act as a whole (see Grano-vetter, 1973; for a review, Csermely, 2006). These proper-ties of the full network – encompassing the contributions of both conceptual and perceptual features – seem highly relevant to some contemporary views of categories – not as fixed partitions – but as functional relations within a system of distributed knowledge (Barsalou, 1999; Samuel-son & Smith, 2000; Tyler, Moss, Durrant-Peatfield, & Levy, 2000). Within such a complex system of connectivity, a horse can be both an animal and a mode of transportation, and the ad hoc category of soft white things can be found and used.

A second potential problem with a framework that seg-regates or privileges conceptual or relational features is that the origins of such relational features themselves are not at all clear (for relevant discussions, see Doumas, Hum-mel, & Sandhofer, 2008; Yoshida & Smith, 2003). Formally, any n-place relation may be redefined as a combination of $n − 1$ place relations, which suggests that functional fea-tures such as "can be worn" might ultimately be under-stood as composed of clusters of interconnected 1-place perceptual features (see Yoshida & Smith, 2003, for a dis-cussion of this idea with respect to animacy features). If this is so, then conceptual features might be not so much fundamentally different from perceptual features but in-stead be themselves dense subnetworks in the larger graph, subnetworks so dense and useful perhaps that lan-guage provides labels for the subnetwork as a whole (e.g., "can be worn") and such that adults then spontaneously offer those labels in feature generation studies. This idea that the nodes of a network are networks themselves are common in graph-theoretic analyses of molecular and cel-lular processes in biology (e.g., Csermely, 2006). Whether these ideas are appropriate to perceptual and conceptual features is not clear at present; what is clear, however, is the perceptual and conceptual features freely offered by adults in feature generation studies contribute in comple-mentary ways to the structure of early-learned categories.

4.3. Testable predictions

If the psychological coherence of higher-order groups is a function of the number of shared features, then the $w = 4$ full network in Fig. 2 presents some intriguing patterns. The subgraphs in this network are composed of categories connected by at least four shared features. By hypothesis, these groupings of containers, vehicles, animals, food, clothing and things to sit on are highly coherent for 2-and-a-half-year-olds. If this category cohesion prediction is true, then in classification tasks, young children should form higher-order categories of these high threshold clus-ters earlier than other clusterings. For example, container is a better superordinate grouping than tools. Furthermore,

the network offers clear predictions about which basic-le-vel categories should be incorporated into these higher-or-der categories. A belt is not well connected to clothing by redundant shared features; a bathtub is not a good con-tainer, and a sled is not a good vehicle.

The graph-theoretic approach taken here also makes predictions about feature generalization. If category cohe-sion predicts category formation, then it may also predict feature generalization as a kind of feature momentum pre-diction: i.e., the probability that two items share one fea-ture is directly proportional to how many other features they share. For example, young children might be expected to generalize some new fact about pants to socks but not to belts, or some new fact about airplanes to tricycles but not to sleds; the latter, in both instances, being less well connected.

In the $w = 4$ full network (and as indicated by the perco-lation clusters), four-legged animals constitute the most densely connected subgraph in the network. As such, four-legged animals should support the most within-kind generalizations, a fact that has been documented in several influential studies of category induction by preschool chil-dren (see Carey, 1985; Gelman & Markman, 1986). Some (e.g., Gelman, 1990; Keil, 1994) have attributed children's seeming precocity in making inferences about animal cat-egories to their evolutionary significance and innate con-ceptual structures. The present results (as do the simulations of Rogers & McClelland, 2004) offer a poten-tially different account based on feature correlation: "four-legged animals" makes a particularly strong group-ing because there are many features that are correlated across four-legged animals.

One can also ask more fine-grained developmental questions about the role of features in category develop-ment. For example, one hypothesis is that children become better able to make category inferences because they be-come better at attending to multiple features, i.e., they can increase w to fine-tune category memberships. Alter-natively, with age may come the ability to selective attend to specific classes of features, e.g., just conceptual features. Finally, one can also ask how features may contribute to learning, by investigating how new noun–feature combi-nations enter the developing network at specific ages according to the MCDI (e.g., Hills, Maoune, Sheya, Maoune, & Smith, 2009).

5. Conclusion

The capacity to create categories from feature correla-tions is a powerful tool for predicting properties about the world. By taking a subset of nouns that many chil-dren know at 30 months and combining these with fea-tures reported to be characteristic of these things, we were able to construct a network that represents a cogni-tive hypothesis about how information is structured in early semantic networks. Analyses of this network re-vealed that it had small-world structure and that the lo-cal structure was consistent with categories that are largely familiar as ad hoc categories of practical utility. We also found clear differences between conceptual and perceptual features. The perceptual network, due to the

abundance and resulting redundancy of perceptual features, maintained local structure under higher thresholds, where as the conceptual network reduced to isolates as the degree of overlap between nouns was increased. Nouns overlapped on several perceptual features but only on a single or very few conceptual features. The pattern of overlap for perceptual features was also such that a given noun could be closely connected to several clusters of densely interconnected nouns that are only sparsely connected to each other. This pattern of overlap allows perceptual features to support several sets of partitions or systems of categories. Whereas conceptual features tend to form isolated collections of densely connected nouns and thus only support a single set of partitions. Both feature types are likely to be important to a functioning category system in which the same information is consistently brought to bear across a variety of contexts, and one in which the information, the set of partitions, is sensitive to changes in context – which set of overlapping perceptual features is relevant could be modulated by the needs of the current context and relevant input from the environment.

Acknowledgements

Appendix. List of nouns

AIRPLANE	CAKE
ALLIGATOR	CARROTS
ANT	CARS
APPLE	CAT
BALL	CHAIR
BALLOON	CHEESE
BANANA	CHICKEN
BASEMENT	CHURCH
BASKET	CLOCK
BATHTUB	CLOSET
BEANS	COAT
BEAR	COMB
BED	CORN
BEDROOM	COUCH
BELT	COW
BENCH	CRAYON
BIKE	CUP
BLACKBIRD	DEER
BOAT	DISH
BOOK	DOG
BOOTS	DOLL
BOTTLE	DONKEY
BOWL	DOOR
BOX	DRESS
BREAD	DUCK
BROOM	ELEPHANT
BRUSH	GOLDFISH
BUCKET	FORK
BUS	FROG
BUTTERFLY	GIRAFFE

GOOSE	PUMPKIN
GRAPE	RADIO
HAMMER	RAISIN
HELICOPTER	ROCK
HORSE	ROOSTER
HOSE	SCARF
HOUSE	SCISSORS
JACKET	SHEEP
JAR	SHIRT
JEANS	SHOES
KEY	SHOVEL
KNIFE	SINK
LAMB	SLED
LAMP	SLIPPERS
LION	SOCKS
MOOSE	SOFA
MOTORCYCLE	SPOON
MOUSE	SQUIRREL
NAPKIN	STICK
NECKLACE	STONE
ORANGE	STOVE
OVEN	STRAWBERRY
OWL	SWEATER
PAJAMAS	TABLE
PANTS	TELEPHONE
PEAS	TIGER
PEN	TRACTOR
PENCIL	TRAIN
PENGUIN	TRAY
PICKLE	TRICYCLE
PIG	TRUCK
PILLOW	TUNA
PLATE	TURKEY
PONY	TURTLE
POTATO	ZEBRA

References

Ahn, W. K., Kalish, C. W., Medin, D. L., & Gelman, S. A. (1995). The role of covariation versus mechanism information in causal attribution. *Cognition, 54*, 299–352.

Ahn, K. W., & Luhman, C. C. (2005). Demystifying theory-based categorization. In L. Gershkoff-Stowe & D. H. Rakison (Eds.), *Building object categories in developmental time* (pp. 277–300). Hillsdale, NJ: Lawrence Erlbaum Associates.

Barabasi, A.-L. (2002). *Linked*. New York: Penguin Group (USA).

Barsalou, L. W. (1983). Ad hoc categories. *Memory and Cognition, 11*, 211–227.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences, 22*, 577–609.

Booth, A. E., & Waxman, S. R. (2002). Word learning is smart: Evidence that conceptual information affects preschoolers' extension of novel words. *Cognition, 84*(1), B11–B22.

Carey, S. (1985). *Conceptual change in childhood*. Cambridge, MA: MIT Press.

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

Csermely, P. (2006). *Weak links, stabilizers of complex systems from proteins to social networks*. Heidelberg: Springer.

De Renzi, E., & Lucchelli, F. (1994). Are semantic systems separately represented in the brain? The case of living category impairment. *Cortex, 30*, 3–25.

Dorogotsev, S. N., & Mendes, J. F. F. (2003). *Evolution of networks: from biological nets to the Internet and WWW.* Oxford, UK: Oxford University Press.

Doumas, A., Hummel, J., & Sandhofer, C. (2008). A theory of the discovery and predication of relational concepts. *Psychological Review, 115,* 1–43.

Edelman, G. M., & Gali, J. A. (2001). Degeneracy and complexity in biological systems. *PNAS, 98*(24), 13763–13768.

Fenson, L., Dale, P. S., Reznick, J., Bates, E., Thal, D. J., & Pethick, S. J. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development, 59*(5), v-173.

Gelman, R. (1990). First principles organize attention to and learning about relevant data: Number and the animate-inanimate distinction as examples. *Cognitive Science, 14*(1), 79–106.

Gelman, S. A., & Bloom, P. (2000). Young children are sensitive to how an object was created when deciding what to name it. *Cognition, 76*(2), 91–103.

Gelman, S. A., & Koenig, M. A. (2003). Theory-based categorization in early childhood. In D. H. Rakison & L. M. Oakes (Eds.), *Early category and concept development: Making sense of the blooming, buzzing confusion* (pp. 330–359). London: Oxford University Press.

Gelman, S. A., & Markman, E. M. (1986). Categories and inductions in young children. *Cognition, 23,* 183–208.

Graham, S. A., & Kilbreath, C. S. (2007). It's a sign of the kind: Gestures and words guide infants' inductive inferences. *Developmental Psychology, 43*(5), 1111–1123.

Granovetter, M. S. (1973). The strength of weak ties. *American Journal of Sociology, 78,* 1360–1380.

Hills, T., Maoune, J., Sheya, A., Maoune, M., & Smith, L. (2009). Longitudinal analysis of early semantic networks: Preferential attachment or preferential acquisition? *Psychological Science, 20,* 729–739.

Holyoak, K. J., & Thagard, P. R. (1995). *Mental leaps: Analogy in creative thought.* Cambridge, MA: MIT Press.

Horton, M. S., & Markman, E. M. (1980). Developmental differences in the acquisition of basic and superordinate categories. *Child Development, 51*(3), 708–719.

Hummel, J. E. (2000). Where view-based theories break down: The role of structure in shape perception and object recognition. In E. Dietrich & A. Markman (Eds.), *Cognitive dynamics: Conceptual change in humans and machines* (pp. 157–185). Mahwah, NJ: Erlbaum Associates.

Keil, F. C. (1979). *Semantic and conceptual development: An ontological perspective.* Cambridge, Mass.: Harvard University Press.

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

Keil, F. C. (1994). The birth and nurturance of concepts by domains: The origins of concepts of living things. In L. Hirschfeld & S. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture* (pp. 234–254). New York, NY, US: Cambridge University Press.

Keil, F. C., & Batterman, N. (1984). A characteristic-to-defining shift in the acquisition of word meaning. *Journal of Verbal Learning and Verbal Behavior, 23,* 221–236.

Kemler Nelson, D. G., Frankenfield, A., Morris, C., & Blair, E. (2000). Young children's use of functional information to categorize artifacts: Three factors that matter. *Cognition, 77,* 133–168.

Kemler Nelson, D. G., Russell, R., Duke, N., & Jones, K. (2000). Two-year-olds will name artifacts by their functions. *Child Development, 71*(5), 1271–1288.

Komatsu, L. (1992). Recent views of conceptual structure. *Psychological Bulletin, 112*(3), 500–526.

Kotovsky, L., & Gentner, D. (1996). Comparison and categorization in the development of relational similarity. *Child Development, 67,* 2797–2822.

Landau, B., Smith, L. B., & Jones, S. S. (1988). The importance of shape in early lexical learning. *Cognitive Development, 3,* 299–321.

Madole, K. L., & Oakes, L. M. (1999). Making sense of infant categorization: Stable processes and changing representations. *Developmental Review, 19*(2), 263–296.

Mandler, J. M. (1992a). How to build a baby II: Conceptual primitives. *Psychological Review, 99*(4), 587–604.

Mandler, J. M. (1992b). The foundations of conceptual thought in infancy. *Cognitive Development, 7*(3), 273–285.

Mandler, J. M., & McDonough, L. (1996). Drinking and driving don't mix: Inductive generalization in infancy. *Cognition, 59*(3), 307–335.

Mareschal, D., Quinn, P. C., & French, R. M. (2002). Asymmetric interference in 3- to 4-month olds' sequential category learning. *Cognitive Science, 26*(3), 377–389.

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

McRae, K., Cree, G. S., Westmacott, R., & De Sa, V. R. (1999). Further evidence for feature correlations in semantic memory. *Canadian Journal of Experimental Psychology. Special Visual Word Recognition, 53*(4), 360–373.

Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review, 92*(3), 289–316.

Nelson, K., & Ware, A. (2002). The reemergence of function. In N. L. Stein, P. J. Bauer, & M. Rabinowitz (Eds.), *Representation, memory, and development: Essays in honor of Jean Mandler* (pp. 161–184). Mahwah, NJ, US: Lawrence Erlbaum Associates.

Palla, G., Barabasi, A.-L., & Vicsek, T. (2007). Quantifying social group evolution. *Nature, 446*(5), 664–667.

Palla, G., Derenyi, I., Farkas, I., & Vicsek, T. (2005). Uncovering the overlapping community structure of complex networks in nature and society. *Nature, 435*(9), 814–818.

Quilian, M. R. (1967). Word concepts: A theory and simulation of some basic semantic capabilities. *Behavioral Science, 12,* 410–430.

Quinn, P. C., & Eimas, P. D. (1996). Perceptual organization and categorization in young infants. In C. Rovee-Collier & L. P. Lipsitt (Eds.). *Advances in infancy research* (Vol. 10, pp. 1–36). Stamford, CT: Ablex Publishing Corp..

Rakison, D. H. (2003). Parts, motion, and the development of the animate–inanimate distinction in infancy. In D. H. Rakison & L. M. Oakes (Eds.), *Early category and concept development: Making sense of the blooming, buzzing confusion* (pp. 159–192). London: Oxford University Press.

Rakison, D. H. (2005). The perceptual to conceptual shift in infancy and early childhood: A surface to deep distinction? In L. Gershkoff-Stowe & D. H. Rakison (Eds.), *Building object categories in developmental time* (pp. 131–158). Hillsdale, NJ: Lawrence Erlbaum Associates.

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

Rosch, E. (1973). Natural categories. *Cognitive Psychology, 4*(3), 328–350.

Rosch, E. (1975). Cognitive reference points. *Cognitive Psychology, 7,* 532–5447.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology, 8*(3), 382–439.

Rumelhart, D. E., & Norman, D. A. (1973). Active semantic networks as a model of human memory. *IJCAI, 1973,* 450–457.

Samuelson, L. K., & Smith, L. B. (2000). Grounding development in cognitive processes. *Child Development, 71*(1), 98–106.

Shapiro, S. C. (1976). *An introduction to SNePS (semantic network processing system).* Technical Report 31. Bloomington, IN: Computer Science Department, Indiana University.

Sheya, A., & Smith, L. B. (2006). Perceptual features and the development of conceptual knowledge. *Journal of Cognition and Development, 7*(4), 455–476.

Sloutsky, V. M., & Fisher, A. V. (2004). Induction and categorization in young children: A similarity-based model. *Journal of Experimental Psychology: General, 133*(2), 166–188.

Smith, L. B. (2005). Emerging ideas about categories. In L. Gershkoff-Stowe & D. H. Rakison (Eds.), *Building object categories in developmental time* (pp. 159–173). Hillsdale, NJ: Lawrence Erlbaum Associates.

Smith, L. B., Jones, S. S., & Landau, B. (1996). Naming in young children: A dumb attentional mechanism? *Cognition, 60*(2), 143–171.

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

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

Tyler, L. K., Moss, H. E., Durrant-Peatfield, M. R., & Levy, J. P. (2000). Conceptual structure and the structure of concepts: A distributed account of category-specific deficits. *Brain and Language, 75*(2), 195–231.

Vitevitch, M. S. (2008). What can graph theory tell us about word learning and lexical retrieval? *Journal of Speech, Language and Hearing Research, 51,* 408–422.

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of small-world networks. *Nature, 393,* 440–442.

Waxman, S. R., & Markow, D. B. (1995). Words as invitations to form categories. *Cognitive Psychology, 29*(3), 257–302.

Yoshida, H., & Smith, L. B. (2003). Correlations, concepts and cross-linguistic differences. *Developmental Science, 6,* 30–34.

Younger, B. A., & Cohen, L. B. (1990). Infants' detection of correlation among feature categories. *Child Development, 61*(3), 614–620.