COGNITIVE SCIENCE
A Multidisciplinary Journal

Feature Biases in Early Word Learning: Network Distinctiveness Predicts Age of Acquisition

Tomas Engelthaler, Thomas T. Hills

Department of Psychology, University of Warwick

Abstract

Do properties of a word's features influence the order of its acquisition in early word learning? Combining the principles of mutual exclusivity and shape bias, the present work takes a network analysis approach to understanding how feature distinctiveness predicts the order of early word learning. Distance networks were built from nouns with edge lengths computed using various distance measures. Feature distinctiveness was computed as a distance measure, showing how far an object in a network is from other objects based on shared and non-shared features. Feature distinctiveness predicted order of acquisition across all measures: Words that were further away from other words in the network space were learned earlier. The best distance measures were based only on non-shared features (object dissimilarity) and did not include shared features (object similarity). This indicates that shared features may play less of a role in early word learning than non-shared features. In addition, the strongest effects were found for visual form and surface features. Cluster analysis further revealed that this effect is a localized effect in the object feature space, where objects' distances from their cluster centroid were inversely correlated with their age of acquisition. Together, these results suggest a role for feature distinctiveness in early word learning.

Keywords: Shape bias; Mutual exclusivity; Network analysis; Word learning; Distinctiveness

1. Introduction

When learning new words, children need to generalize word-object-mappings across different items that vary along numerous dimensions. For example, suppose a child learns the word "spoon" in relation to one spoon. In learning new words for other objects, they need to be able to distinguish the category of spoons (which do not need new labels) from other categories of objects that do, such as forks, bowls, and toothbrushes. If an

Correspondence should be sent to Tomas Engelthaler, Department of Psychology, University of Warwick, Gibbet Hill Road, CV4 7AL, Coventry UK. E-mail: t.engelthaler@warwick.ac.uk

item looks too much like a spoon (e.g., a spork), the inability to distinguish this item from other spoons may, in principle, make it more difficult to learn the new word. Children are proposed to overcome this apparently difficult task with the assistance of learning biases (Markman, 1990). In this article, we will combine what we see as two complimentary biases that help children solve this problem, and we will use these to develop and investigate a new notion of feature distinctiveness in age of acquisition.

One bias known to influence word generalization is *shape bias*. When children are asked to learn the name of an unfamiliar object, they tend to generalize the word to other objects based on shape rather than other feature types, such as texture, color, or material (Landau, Smith, & Jones, 1988). For example, if a child is presented with a round, rubber object in association with the word "dax," it is more likely that the child will generalize the word "dax" to other objects that are round, rather than generalizing "dax" to other objects made of rubber. Shape bias is not always found when investigating early world learning (Cimpian, & Markman, 2005) and is clearly influenced by developmental trends (see Landau et al., 1988). Questions about the origination and generality of shape bias are quite common in the literature (Booth & Waxman, 2008; Kemp, Perfors, & Tenenbaum, 2007; Markson, Diesendruck, & Bloom, 2008), as well as how shape bias may be influenced by more general learning or feature biases. These are the topics we take up here.

There are a number of additional biases that may be considered to fall into the broad class of *feature biases*, with the two most prominent being texture and material biases (e.g., Jones, Smith, & Landau, 1991; Samuelson & Horst, 2007; Soja, 1992; Soja, Carey, & Spelke, 1991). In addition, there are a number of word features associated with phonology and structure in the language (Hills, 2013; Iversen, Patel, & Ohgushi, 2008; Morgan, 1996; Thiessen & Saffran, 2007). Here, we are only concerned with the early influence of object feature biases by which children generalize word labels associated with objects to different degrees depending on their feature similarity. Shape bias would then represent situations where the unequal generalization of word labels is correlated with a difference in shape, but not a difference in other feature dimensions.

A second bias known to influence word learning, which may work in tandem with feature biases, is *mutual exclusivity*. In early lexical development children prefer to assign novel words to objects that do not yet have names. This means that two different words are often interpreted to refer to two different objects (Markman, Wasow, & Hansen, 2003; but see Bilson, Yoshida, Tran, Woods, & Hills, 2015). Mutual exclusivity has been shown to influence word-object associations and this is related to a preference for labeling novel objects with novel words (e.g., Mather & Plunkett, 2009, 2012). In turn, mutual exclusivity is likely to play a strong role in early word learning (Hills, 2013; Kachergis, Yu, & Shiffrin, 2012; Yurovsky, Yu, & Smith, 2013).

Mutual exclusivity may play a still broader role in learning if it extends beyond individual objects to categories. If children can assess whether a newly encountered object is a member of an already named category, then they can assign novel names to objects in novel categories: If an object belongs to an already named category, it should be less likely to receive a new label. How categories are defined should be influenced by chil-

dren's perceptual and cognitive systems and what properties or features of objects children pay attention to (Colunga & Smith, 2008; Smith & Samuelson, 2006). For example, the more visually dissimilar objects are, the more inclined children are to show a shape bias (Tek, Jaffery, Swensen, Fein, & Naigles, 2012). Feature distributions over categories are also a central concept in the categorization literature, where the influence of feature types and shared versus distinctive features is well known to influence category learning (Love, Medin, & Gureckis, 2004; Sloutsky & Fisher, 2004; Weitnauer, Carvalho, Goldstone, & Ritter, 2014). If children use mutual exclusivity at the category level, they must do so based on information provided in the shared and non-shared features between objects. This sensitivity to the feature similarity between objects, in turn, may facilitate feature biases in the process of lexical acquisition.[1]

The majority of research on shape bias has been experimental in nature (e.g., Booth, Waxman, & Huang, 2005; Collisson, Grela, Spaulding, Rueckl, & Magnuson, 2015; Graham & Diesendruck, 2010). This research strongly supports a link between shape bias and word learning. Specifically, in a study by Smith and colleagues, a group of children too young to systematically use shape bias were trained over a 9-week period by exposure to novel words for novel objects, presented in categories organized by shape. Children with exposure to this training learned nouns outside the laboratory at a faster rate than control children. This suggests some sensitivity to shape-based (and possibly other feature-based) categories during word learning (Smith, Jones, Landau, Gershkoff-Stowe, & Samuelson, 2002).

If mutual exclusivity and feature biases work together to facilitate early lexical learning, then structural relations between one object's features and the features of other objects a child is exposed to should provide information with respect to an object's age of acquisition. In this study, we look at some of the implications of combining feature biases and mutual exclusivity. Through feature network analysis we aim to offer insight into how these implications are reflected in age of acquisition.

To date, however, extensions of feature information to predict age of acquisition have been unsuccessful. Hills, Maouene, Maouene, Sheya, and Smith (2009a) investigated the role of feature similarity in age of acquisition by producing weighted networks of words based on shared features. In these networks, nodes in the network represented nouns and edge weights between nodes represented the number of shared features. Using several different modeling approaches, Hills et al. (2009a) found no evidence that feature similarity could predict age of acquisition. While subsequent work has investigated the role of semantic and associative factors in early word learning (Hills, Maouene, Riordan, & Smith, 2010; Hills, 2012), to our knowledge no further progress has been made with features. However, one potential problem with the approach in Hills et al. (2009a) is that it overlooks the inference from mutual exclusivity outlined above—that is, learning may be predicted not by shared features but by feature distinctiveness, which we operationalize here as a measure of non-shared features between one or more objects.

Network analysis, or graph theory, allows investigations of structured information based on the relationships between objects or nodes. In studies of language learning and language processing, the relationships (i.e., edges) between nodes have been based on, for

example, phonetic similarity (e.g., Arbesman, Strogatz, & Vitevitch, 2010), free associa-
tion norms (e.g., Hills et al., 2009a), and co-occurrence in language (e.g., Beckage,
Smith, & Hills, 2011; Hills et al., 2010). In the present work, we use several quantitative
measures of distinctiveness to produce networks of words for objects based on the fea-
tures of those objects. In these networks, edges now represent how dissimilar are the ref-
erents of two words. This then allows us to compute distinctiveness for objects in the
network, as a function of their overall dissimilarity to other objects in the network. In
addition, this also allows us to investigate additional structural properties of distinctive-
ness, such as to what extent this effect is driven by specific feature types or is a local
(near neighbor) or global (across all words) property.

Our approach is based on the inference that feature biases and mutual exclusivity
should lead to two distinct patterns in age of acquisition. First, words representing objects
that are more distinctive should be learned earlier than words for objects that are less dis-
tinctive. That is, if the basis for two objects being considered the same is a function of
their shared and non-shared feature distribution, then objects that share fewer features
with other objects—which are therefore more distinctive—should be learned earlier.
Mutual exclusivity could then be viewed as an example of a distinctiveness-driven rela-
tionship between objects, rather than a stand-alone learning bias. Second, the classic
account of shape bias (as a prominent subset of feature biases) should predict that the
above finding will be more prominent with features relating to object's shape, rather than
other non-shape-related features of an object (e.g., function, sound, or material). As noted
above, the precedence of shape over other feature categories (e.g., function) has been sub-
ject to conflicting experimental results (e.g., Diesendruck & Bloom, 2003). However, it
may also be that shape bias is a component of feature biases, where each of the compo-
nent biases (including texture and material) is driven by a similar process of distinctive-
ness.

Finally, we also ask to what extent distinctiveness represents a global or local property
by investigating how distinctiveness operates within categories (i.e., subnetworks) in com-
parison with all words simultaneously. We know of no prior work on the topic of local
versus global distinctiveness. However, for distinctiveness to be effective for word learn-
ing it is particularly important that it function locally, allowing children to discern similar
objects from one another. Should distinctiveness only work on a global level, it may be
primarily driven by the fact that some superordinate categories of words may overall be
more distinctive than other superordinate categories. Such an effect would be less useful
when trying to distinguish two closely related objects. The contribution of local effects is
investigated using network clustering analyses.

Our goal is to better understand how feature distinctiveness might contribute to
early learning biases (e.g., shape bias). Specifically, if feature distinctiveness is driving
the early influence of processes like mutual exclusivity, it may also provide a
broader explanation for feature biases more generally, and offer an explanation for
previous findings in the literature. Word learning is a complex process shown to
involve a large variety of factors. The present work is not an attempt to explain how
word learning works as a whole. Instead, it is aimed to be a novel approach to

understanding how feature distinctiveness might affect certain feature-related learning biases.

2. Methods

2.1. Features

We used the McRae semantic feature production norms as the basis of our network generation (McRae, Cree, Seidenberg, & McNorgan, 2005). This is a collection of 541 living and non-living concepts (nouns) with features collected from approximately 725 adult participants for each concept. These include 7,259 unique features, which are labeled and categorized both according to the division proposed by Cree and McRae (2003) and also the taxonomy of Wu and Barsalou (2009). Each listed feature is assigned to one Cree & McRae category and one Wu & Barsalou category. See Table 1 for examples of the listed features and their Cree & McRae category assignment.

These feature norms make it is possible to see each object as a concept representing a list of features. Furthermore, the division of these features into categories allows us to quantify to what extent objects differ from one another in terms of these feature categories (e.g., *Does one word represent a higher proportion of visual features than another?*).

Some developmental studies indicate that children do not simply learn words, but also perceive them as categories containing features (Sheya & Smith, 2006). While children may or may not view the world through the fine-grained categories listed above, it has been established that children can and do use object features as one of the main components of lexical development and that they do use broader categories that with age become similar to adult categories (Keil & Batterman, 1984; Schyns, Goldstone, &

Table 1
Examples of listed features with their respective assigned categories

Assigned Category (Exhaustive List)	Three Randomly Sampled Example Features
Visual form and surface	Has legs, is big, made of wood
Visual motion	Runs, crawls, is fast
Visual color	Is green, is dark, is colorful
Taxonomic	Is a fruit, is an animal, is a tool
Encyclopedic	Used long ago, found in houses, made by bees
Function	Is eaten, used for building, requires slicing
Sound	Is loud, is buzzing, plays music
Tactile	Is rough, is sharp, is soft
Taste	Is sweet, tastes good, tastes hot
Smell	Is smelly, smells nice, smells bad

Notes. Each of the 7,259 listed features is assigned to exactly one of the 10 Cree & McRae categories. The "assigned category" is an exhaustive list of all the possible categories. The three example features have been randomly sampled to illustrate what features these categories contain.

Thibaut, 1998). Infants also make links between words, their referents, and the referents' functions to facilitate lexical learning, hinting at the ability to understand deeper connections behind words (Booth & Waxman, 2002). At the same time, the extent to which children understand deeper categories is a topic of continuing discussion within the literature. Some studies suggest children's learning of categories may not mean children understand the deeper structure of the category (Plunkett, Hu, & Cohen, 2008). The "encyclopedic" feature category may be an especially inaccurate representation of how a child views the world, as it relies on extensive general knowledge. For that reason, we removed encyclopedic features from our analyses, reducing the dataset to 5,842 features. Leaving them in, however, does not change the general conclusions of this work, with similar results throughout. With this in mind, we encourage the reader to not interpret the feature categories as direct representations of a child's view. Instead, we suggest this study is an exploratory view into what adult-perceived word categories are most susceptible to distinctiveness effects in relation to age of acquisition.[2]

2.2. Age of acquisition

We obtained the age of acquisition of each word used in this study from the Kuperman norms dataset (Kuperman, Stadthagen-Gonzalez, & Brysbaert, 2012). This is a database of over 30,000 words, ranked on age of acquisition by a large sample of online participants. These norms show the retrospectively estimated age of acquisition of a given word. This database also contains 492 words in the McRae feature dataset (out of 541), which we used in this study. The age of acquisition rating represents the year in which an adult participant estimated that they understood a given word. The data used in our study is the mean age of acquisition rating for a given word—this is the mean retrospective age of acquisition rating based on 20 participants' responses to each word. It is important to note that these are retrospective estimates of age of acquisition by an adult population. Adult age of acquisition ratings are generally much easier to collect than parent-observed (direct) age of acquisition indexes, resulting in bigger datasets. The bigger dataset is the primary reason for opting to focus on adult retrospective ratings in this study. Using the Kuperman norms, our pool of analyzable data grows considerably in contrast to using parent-indexed age of acquisition. One can get some idea of how well the Kuperman (adult retrospective) norms actually relate to children's lexical pool by correlating the ratings with a parent-indexed age of acquisition rating dataset. One such parent-indexed dataset is the MacArthur-Bates Communicative Developmental Inventory, toddler version (MCDI; Dale & Fenson, 1996). This is a set of 680 words along with production metrics for 1,789 children, collected monthly from caregivers of children between the ages of 16 to 30 months. These show the proportion of toddlers using a given word at a specific month of age. For the words that were present in both norm databases, the reported age of acquisition in the Kuperman norms was positively correlated with the reported MCDI age of acquisition (first recorded month of age where at least 50% of the toddlers demonstrated the use of a given word) $r_s(107) = .64$, $p < .001$.[3] While the Kuperman norms correlate well with direct age of acquisition measures and also with a cross-validation

adult sample (see Kuperman et al., 2012), our main results reflect retrospective adult ratings. Inferences to child learning are based on the assumption that retrospective adult age of acquisition ratings are representative of the size of a child's lexical pool. While we believe this to be true, additional possible explanations are also considered in the discussion.

Furthermore, the mean adult-rated estimate age of acquisition for our sample of 492 words was 4.3 years ($SD = 0.83$). The mean parent-observed (direct) age of acquisition for the same word sample, collected from the MacArthur-Bates Communicative Developmental Inventory, toddler version (MCDI; Dale & Fenson, 1996), was 1.8 years ($SD = 0.28$). Combined with the mean rating correlation mentioned in the previous paragraph, this suggests that while adult retrospective ratings may be generally indicative of which words are learned earlier in relation to other words, the rating may be inaccurate with respect the exact time of learning. Specifically, all our results should be viewed as based on relative age-of-acquisition relationships between words, not claims about how distinctiveness relates to absolute age of acquisition (e.g., specific months of learning).

For the above reasons, the majority of our analyses are non-parametric, based on rank measures rather than absolute values. This is not only because of normality assumptions, but also to ensure the effects are not misinterpreted due to the inconsistency in absolute age between adult age of acquisition ratings and parent-observed indexes.

2.3. Distinctiveness measures

Our measures of distinctiveness involve both counts of distinctive features as well as network representations. *Relative feature distinctiveness* is a measure of how rare a noun's individual features are with respect to all other words in the norms. This is defined as the sum of the distinctiveness of all the features reported for a word:

$$x = \sum_{i=1}^{m} \frac{1}{w_i} \tag{1}$$

where x is the overall relative feature distinctiveness of a word; m is the total number of features listed for a word; and w_i is the number of words listing feature i. Thus, the more objects that have a feature, the less distinctive the feature. Note that this measure is not computed at the pair-wise level between words, but computes rareness over features and then sums this for individual words.

2.3.1. Network measures

Our network distinctiveness measures involve creating a node for each noun in the McRae feature dataset. The distance between nodes (the edge length) in the network is a measure of relative feature distinctiveness between the two nodes computed as described below. The categorization of feature types mentioned above allowed us to generate networks based on all features and also based on subsets of features associated with specific

categories, where the distances between the nodes of a network then represent feature dissimilarity on only one feature dimension (e.g., "visual color").

We calculated the network distinctiveness using three different approaches, which each make different underlying assumptions about the features involved. These are the non-shared feature distance, the Jaccard distance, and the Manhattan distance. Together, these three measures allow us to isolate the effects of shared features (comparing Jaccard distance with non-shared feature distance) and the role of saliency (comparing Manhattan distance with non-shared feature distance).

The *non-shared feature distance* calculates the distance between two nouns as the sum of all the features the two respective words do not have in common. In set theory, this is known as the symmetric difference between the two feature sets:

$$d = (n_1 + n_2) - 2n_s \tag{2}$$

where n_1 is the total number of features listed for the first word; n_2 is the total number of features listed for the second word; and n_s is the total number of features that were shared by word one and two. This measure focuses only on the dissimilarity between two concepts, because shared features are excluded. It also represents a feature listed by one respondent with the same strength as a feature listed by ten respondents, and thus does not weight features in relation to the frequency with which they are produced.

The *Jaccard distance* calculates the distance between two nouns as the ratio of the symmetric distance between the intersection of the two nouns' feature sets and their union. Thus, the Jaccard distance normalizes the distance in relation to the total number of features available for comparison. Formally, the Jaccard distance between two nodes is

$$d = 1 - \frac{n_s}{n_1 + n_2 - n_s} \tag{3}$$

where n_1 is the total number of features listed for the first word; n_2 is the total number of features listed for the second word; and n_s is the total number of features that were shared by word one and two.

In contrast to all our other measures we use, the Jaccard distance is sensitive to shared features. Nouns with the same number of non-shared features but with a larger set of shared features will have a smaller Jaccard distance than nouns with fewer shared features. Thus, Jaccard distance is reduced when shared features are added.

The *Manhattan distance* calculates the distance between two nouns as the sum of the absolute differences between the proportions of participants reporting each feature. Thus, the Manhattan distance measure takes into account feature salience, where salience is indicated by the number of participants who produce a given feature. The Manhattan distance is

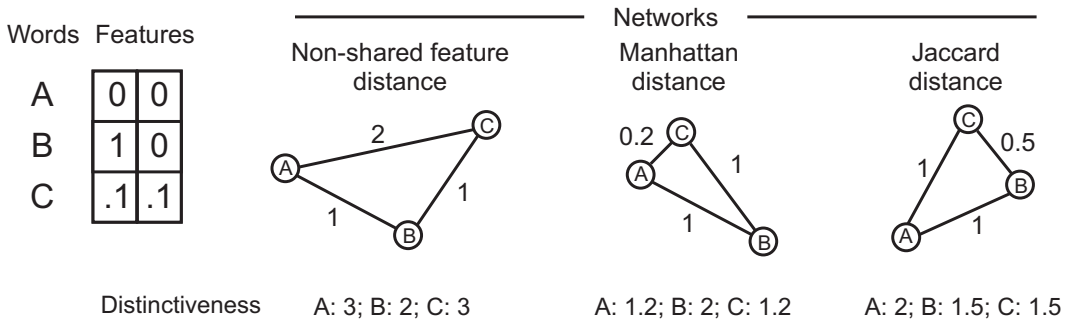$$d = \sum_{i=1}^{n} |p_i - q_i| \tag{4}$$

Fig. 1. An example of the three different distance measures used to compute feature distinctiveness. Each of the three words (A, B, and C) has a vector representing the proportion of individuals who reported each of two features; each feature is represented by a column in the feature matrix shown on the left. To the right, networks are shown for each of the distance metrics described in the text. Distinctiveness for each word is provided beneath the networks.[4]

where n is the number of features recalled for both words (i.e., the union of the two feature sets); p is the proportion of people listing feature i for the first word and q is the proportion of people listing feature i for the second word. Proportion data can be found in the McRae norms and is defined as the total number of people listing a feature for a given word divided by 30 (each word was annotated with features by 30 people). Unlike the other measures, Manhattan distance places emphasis not on the difference in number and types of features listed, but quantifies to what extent these differences in features are salient. Figure 1 shows an example of how these three measures, provided the same input data, generate different networks.

3. Results

3.1. Relative feature distinctiveness predicts age of acquisition

Overall relative feature distinctiveness was negatively correlated with Kuperman's age of acquisition $r_s(490) = -.21$, $p < .001$,[5] showing that words with more distinctive features are perceived to be learned earlier. This negative correlation was also found between overall distinctness and the MCDI (parent-reported) age of acquisition, $r_s(111) = -.19$, $p < .05$, showing that words with more distinctive features are learned earlier when using the parent-reported age of acquisition norms as the base of the analysis.[6]

3.2. Relative feature distinctiveness of visual form and surface features best predicts age of acquisition

The division of features into feature types allows us to investigate whether features of different types predict age of acquisition to differing extents. Each feature was labeled

Table 2
Variance for each feature category as an outcome of a linear regression model

Variable	Age of Acquisition	
	B	CI
Visual form and surface	−0.12*	[−0.20, −0.03]
Visual motion	−0.10*	[−0.19, −0.00]
Sound	−0.09	[−0.18, 0.00]
Tactile	−0.06	[−0.15, 0.03]
Function	−0.03	[−0.12, 0.06]
Taste	−0.02	[−0.11, 0.07]
Smell	−0.02	[−0.11, 0.07]
Visual color	−0.00	[−0.09, 0.09]
Taxonomic	−0.00	[−0.09, 0.10]

Notes. Relative feature distinctiveness predicting adult reported age of acquisition. $N = 492$. Adjusted $R^2 = .02$, $F = 2.26*$. B = standardized beta coefficient. CI = 95% confidence interval. *$p < .05$.

with one of ten types as proposed by Cree and McRae (2003)—for example, "visual form and surface," "visual motion," etc. For each word, we computed the relative feature distinctiveness for each of the feature types for which it contained features. We then used the relative feature distinctiveness across feature types to predict age of acquisition. A multiple linear regression revealed that only two feature types were predictive of age of acquisition—"visual form and surface" and "visual motion" (Table 2). Visual form and surface was the best predictor among all the features.[7]

3.3. Network distinctiveness predicts age of acquisition

As shown in Table 3, both the non-shared feature network and Manhattan distance network showed negative correlations between distinctiveness and age of acquisition— $r_s(490) = −.24$, $p < .001$ and $r_s(490) = −.34$, $p < .001$, respectively. The Jaccard distance network did not show a significant correlation ($r_s(490) = −.04$, $p > .05$). These correlations show that network measures of distinctiveness based on dissimilarity are more strongly correlated with age of acquisition than a measure of relative feature

Table 3
Word distinctiveness and age of acquisition

Model	r_s	CI
Non-shared feature distance	−.19***	[−0.27, −0.11]
Jaccard distance	−.03	[−0.12, 0.06]
Manhattan distance	−.28***	[−0.36, −0.20]

Notes. $N = 492$. $df = 490$. r_s = Spearman rank-order coefficient. CI = 95% confidence interval. ***$p < .001$.

Table 4
Regression table for distinctiveness measures predicting Kuperman age of acquisition

Variable	Age of Acquisition	
	B	CI
Non-shared feature distance	−0.03	[−0.20, 0.13]
Jaccard distance	−0.06	[−0.14, 0.03]
Manhattan distance	−0.17*	[−0.33, −0.01]

Notes. $N = 492$. Adjusted $R^2 = .04$, $F = 7.07$***. B = standardized beta coefficient. CI = 95% confidence interval. *$p < .05$.

distinctiveness, and they are also consistent with our prediction based on feature biases and mutual exclusivity.

When the three distance measures were analyzed as predictors in a linear regression model, only Manhattan distance explained additional variance after controlling for the other two distance measures (Table 4).

As mentioned in the methods section, the Jaccard distance metric is sensitive to shared features. On the other hand, non-shared feature distance and Manhattan distance use dissimilarity in feature makeup to calculate distance between words. The absence of a significant correlation of Jaccard distance suggests that presence of shared features does not relate to age of acquisition the same way the presence of distinctive features does. Hence, feature dissimilarity (rather than feature similarity) may be more important when looking at age of acquisition effects. This may also explain the limited predictive power of the network similarity approach taken in Hills et al. (2009a), which was based exclusively on shared features.

3.4. Distinctiveness within feature types

As noted above for relative feature distinctiveness, some feature types may be more salient than others and, in turn, more correlated with age of acquisition. The developmental progression of feature salience may be taken to indicate that non-perceptual features should be less predictive than perceptual features (e.g., Sloutsky, 2010). To address this, we used Manhattan distance—our most predictive distinctiveness measure—to construct networks for each feature category separately. This allowed us to calculate the overall per-category distance for each word. The per-category distance was computed by running our Manhattan distance analysis on only a subset of the features—in other words, a metric of how distinct of a feature makeup a word has only taking into account one feature category at a time. This measure then shows how distinctiveness by category correlates with age of acquisition, resulting in a correlation table of category-specific distinctiveness by age of acquisition. Table 5 shows that the strongest correlation was with "visual form and surface" distinctiveness, showing a negative correlation with age of acquisition r (484) = −.19, $p < .001$. This finding suggests that words that are less similar to other words in a "visual form and surface" network are learned earlier. On the other hand, the "taxonomic" category is also significant, although in the opposite direction, $r(394) = .10$,

Table 5
Relationship between feature network word distance and age of acquisition

Model Feature Type	r_s	CI	n
Visual form and surface	−0.19***[bbb]	[−0.28, −0.12]	486
Smell	−0.18	[−0.56, 0.25]	19
Taste	−0.14	[−0.39, 0.13]	51
Tactile	−0.11	[−0.28, 0.05]	165
Visual color	−0.08	[−0.20, 0.04]	264
Sound	−0.05	[−0.29, 0.17]	93
Visual motion	−0.05	[−0.21, 0.11]	165
Function	0.05	[−0.05, 0.15]	418
Taxonomic	0.11*	[0.01, 0.21]	396

Notes. n = number of observations. df = (*n*−2). *r_s* = Spearman rank-order coefficient. CI = 95% confidence interval. **p* < .05. ****p* < .001. [bbb] = *p* < .001. (Bonferroni correction).

$p < .05$. Here, the inference may be that more taxonomically distinctive words are learned later, indicating they may belong to more uncommon taxonomic categories.[8]

3.5. Distinctiveness networks within "visual form and surface"

The "visual form and surface" feature category is the largest feature category, covering 32% of all the reported features. This category can be further broken down into more specific subsets. To do this, we used the taxonomy proposed by Wu and Barsalou (see McRae et al., 2005, for information on the taxonomy) to investigate the distinctiveness within visual form and surface features. This resulted in subsetting visual form and surface features into five subcategories (omitting two additional categories that represented less than five features). We then constructed Manhattan distance networks for each word in each subcategory network and calculated the overall network distance for each word. Table 6 shows that only two groups displayed significant negative correlations with age of acquisition: "external surface property" $r_s(344) = −.12, p < .05$; and "material" $r_s(238) = −.25, p < .001$. These findings suggest that only two of the visual form and surface subgroups are indicative of distinctiveness associated with age of acquisition, where words for more distinctive items are learned earlier. We address this further in the discussion in relation to shape bias.

3.6. Global versus local feature distinctiveness as a predictor

All the above distinctiveness measures are global. That is, feature distinctiveness is measured as a function of relationships between *all* nouns in the dataset. However, this does not allow us to distinguish between local and global distinctiveness, and it may indicate that object names are learned either (a) because they are in distinctive clusters (small clusters of objects that are fairly distinct from all other objects) or (b) because they are fairly distinct within a cluster. In laboratory studies distinctiveness is implicitly taken to be a local measure among other objects in the study, though some objects may be more or less similar to other items the child is familiar with. In previous work, Hills, Maouene,

Table 6
Correlation between Manhattan distance of words and age of acquisition for various feature types within visual form and surface networks

Model Feature Type	r_s	CI	n
Made of (material)	−0.25***[bbb]	[−0.36, −0.13]	240
External surface property	−0.12*	[−0.22, −0.02]	346
Internal surface property	−0.10	[−0.50, 0.31]	29
External component	0.00	[−0.10, 0.11]	377
Internal component	0.12	[−0.06, 0.28]	113

Notes. n = number of observations. df = $(n−2)$. r_s = Spearman rank-order coefficient. CI = 95% confidence interval. *$p < .05$. ***$p < .001$. bbb = $p < .001$. (Bonferroni correction).

Maouene, Sheya, and Smith (2009b) found that objects clustered by features tended to form meaningful categories, such as clothes, food, and animals. Moreover, feature similarity strongly influences children's inductive generalizations (Sloutsky, 2010; Sloutsky & Fisher, 2011). For object distinctiveness to be useful for word learning, distinctiveness within categories should also be predictive of age of acquisition.

To test this, we used the spherical k-means network clustering algorithm to cluster nouns in relation to feature similarity. Spherical k-means finds an appropriate clustering given a number of clusters, k, such that objects are placed in clusters with the nearest mean similarity (Dhillon & Modha, 2001). In the present case, words in the same cluster are likely to share similar features, based on the cosine similarity of their shared feature vectors. Following clustering, we computed the mean Manhattan distance for each noun from its other category members (i.e., the distance to the cluster centroid). For each cluster, we then correlate age of acquisition with cluster centroid distance. We computed the analyses for a range of cluster numbers (5 to 50). Figure 2A shows the mean average correlation for all clusters within a given cluster breakdown. Similarly, Fig. 2B shows the proportion of clusters showing a significant p value ($p < .05$) This allows us to see both what number of clusters shows the strongest mean effect, but also highlights the fact that our results are apparent over a range of assumptions about the number of clusters.

Table 7 shows the cluster breakdown for the cluster number $k = 7$, which had the largest proportion of clusters, showing a significant correlation between centroid distance and age of acquisition (71%). As shown, a noun's feature distance from its category members is strongly negatively correlated with age of acquisition across the majority of clusters. Words most distinctive from their other category members are learned earliest. Distinctiveness therefore appears to function as a local measure of how dissimilar are the most similar objects, and would therefore be appropriate for word learning within these subcategories.

4. Discussion

Object features influence how we experience the world and how we discriminate and name different objects. In the present work, we showed that it is viable to explore
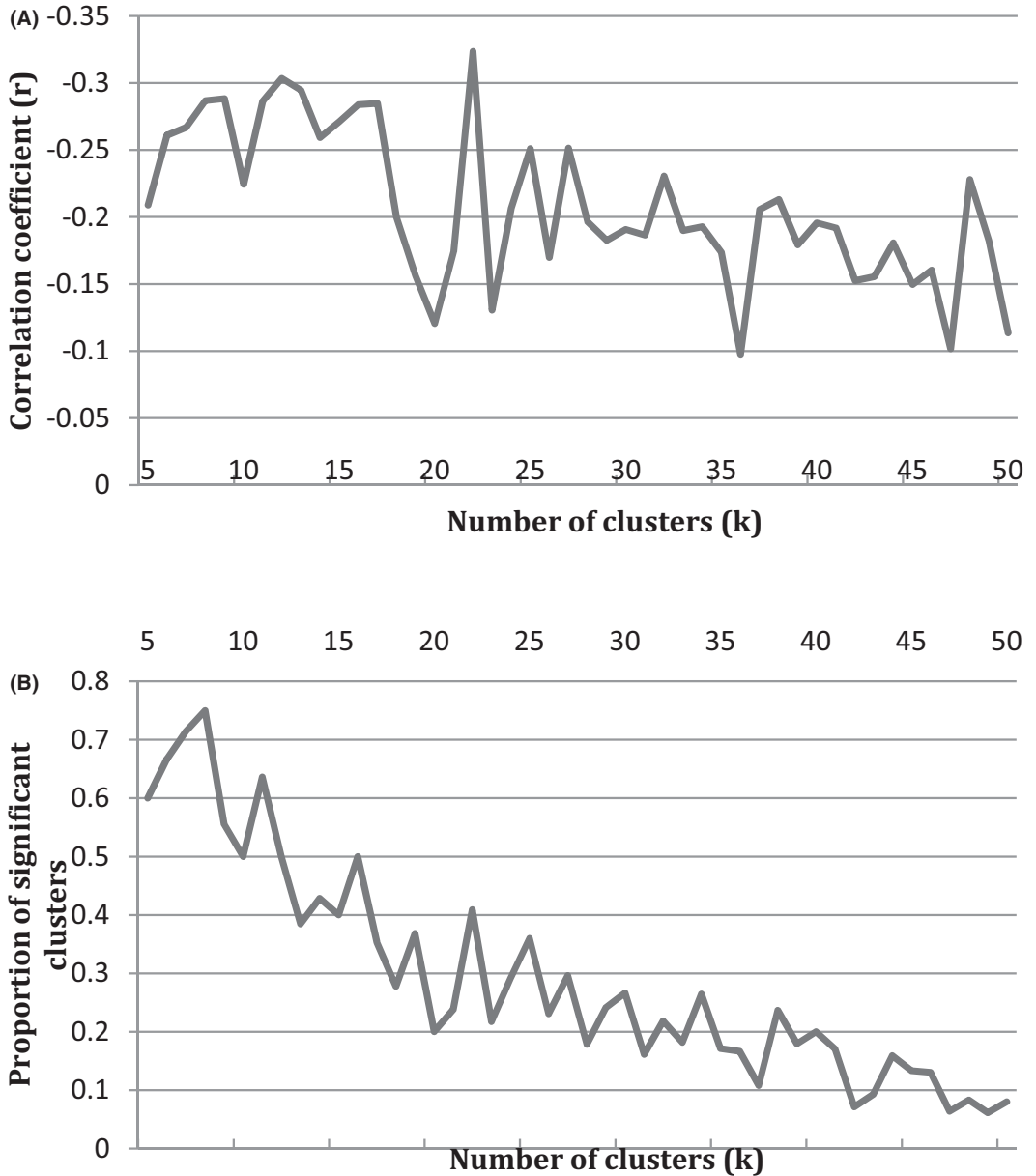
Fig. 2. (A and B) Mean correlation coefficients and proportionate *p* value significance of within-cluster distinctiveness and age of acquisition. For each cluster number, *k*, words are divided into *k* clusters using spherical-*k*-means, then the words' Manhattan distance from all other members of its cluster were computed and correlated with age of acquisition. One correlation is calculated per cluster, with the mean across all clusters presented in (A). The proportion of clusters showing a significant *p* value in said correlation is presented in (B).

Table 7
Correlation between words' distinctness within a cluster and age of acquisition

Cluster	r	CI	n	Exemplars
1	−0.21	[−0.43, 0.05]	61	Car, trolley, piano
2	−0.34*	[−0.55, −0.08]	55	Skyscraper, door, box
3	−0.31**	[−0.48, −0.12]	98	Fork, sink, hatchet
4	−0.25*	[−0.43, −0.05]	97	Broccoli, honeydew, sardine
5	−0.36*	[−0.58, −0.09]	49	Butterfly, swan, wasp
6	−0.35**	[−0.55, −0.13]	68	Mouse, coyote, sheep
7	−0.29*	[−0.50, −0.05]	64	Scarf, wand, necklace

Notes. n = number of observations. df = $(n-2)$. r = Pearson product-moment correlation coefficient. CI = 95% confidence interval. Exemplars show three example members of a given cluster. Spherical k-means clustering algorithm was used to categorize data into seven clusters. *$p < .05$. **$p < .01$.

feature biases and mutual exclusivity through network analysis, and that this approach can elucidate potential factors influencing early word learning. The present work makes three contributions to this area of research. First, object distinctiveness was negatively correlated with age of acquisition, implying that words associated with more distinctive object features are learned earlier. Second, different feature types contribute to this effect to differing extents, with the principal feature types associated with visual form and surface properties. Third, using cluster analysis, we demonstrated that this effect was a local property of distinctiveness, indicating that these effects may come about via distinctiveness between near neighbors in the feature space as opposed to items belonging to distinctive categories. In what follows we briefly describe these results and their implications.

Throughout the literature, mutual exclusivity has been defined as the tendency to pair one object with only one word label (Jaswal & Hansen, 2006; Markman & Wachtel, 1988). This concept is often framed as a word learning principle. However, if word learning is in fact an example of a categorization process, then mutual exclusivity may represent an example of a principle based on categorization more generally. As such, mutual exclusivity may be framed as a property of categorization with the implication that the more dissimilar two objects are, the less likely they are to be assigned to the same category. In the context of word learning, this would mean that dissimilar objects are more likely to take on a new label—a finding demonstrated by the present paper. This view—of mutual exclusivity as a categorization facilitator—is a novel addition to how the field has framed mutual exclusivity so far, and one we feel is supported by the present research.

All our analyses showed that distinctiveness negatively correlates with age of acquisition—meaning that more distinct words are learned earlier. This finding offers additional support to the mutual exclusivity principle, which we interpreted to suggest that children should learn names for more dissimilar objects more easily. The observation that Manhattan distance and non-shared feature networks are correlated with age of acquisition, whereas the Jaccard distance is not, suggests that mutual exclusivity may be based more

on distinguishing features of objects than on the number of shared features. The role played by shared and distinct features in mutual exclusivity is an important line of future research.

The importance of distinctive versus shared features was demonstrated by the relative performance of Manhattan distance versus Jaccard networks. The accuracy of the Jaccard metric relies on comparing the proportion of non-shared features to overall features in the union of the two features sets for two objects. On the other hand, the Manhattan distance calculation is built around contrasting two concepts on differences in feature salience (number of participants producing the feature); the addition of equally salient shared features does not influence this calculation. This suggests, perhaps paradoxically, that the kinds of information that make items similar (e.g., in relation to inductive inference) may be quite different from the kinds of information that make objects distinctive (e.g., in relation to the mutual exclusivity principle). This suggests a cognitive model that places different weights on shared versus distinctive information—a finding that may further inform the underlying assumptions behind shape bias (Gentner & Imai, 1995; Samuelson & Smith, 1999).

Investigating different dimension weighting in more detail, our regression showed that visual form and surface features were the best predictor of age of acquisition among features. Looking at visual form and surface more specifically, we found that external surface property and material are particularly important. External surface property covers features like "is big" and "is round"—features that define the general shape of an object. Material defines the material of an object such as, for example, "water" or "rubber." Additional features also explained variance in age of acquisition in Table 2. Visual motion tells us how something moves, "it flies," and tactile represents how it feels to the touch, "is soft." These findings imply that the notion of shape bias is not as simple as comparing the contours of two objects and claiming they are similar if they match and dissimilar if they differ. Instead, it seems plausible that children make an assessment based on general shape in combination with other features (see Samuelson & Smith, 1999). The classic shape bias experiments that juxtapose contour and texture comparison may be looking at an important subsection of a higher dimensional comparison process.

The low predictive power of external components (features such as "has legs" and "has buttons") has several potential explanations. First, it is possible that the nature of the feature coding process results in features that are overly specific or, alternatively, overly generalized. Our results depend on the processes used by adults to generate features, and this is very likely to contain biases that influence our results. For examples, defining a squid as "has tentacles" may not be particularly useful if few other objects share this property and many other objects have numerous distinctive external components as well. On the other hand, defining a human as "has legs" may be too general, as many animals have legs. Thus, it may be the case that external components are not sufficiently detailed in relation to shape to predict age of acquisition—representing a kind of ceiling effect created by obtuseness of the data. Moreover, if each object is discriminable by a specific variation in external components, then the distinctiveness of

external components may be uninformative, as they are all distinct. This explanation is supported by research on visual perception, as most of the visual perception studies agree that correct recognition of an object relies on successfully identifying subelements of the whole picture (Biederman, 1987; Logothetis & Sheinberg, 1996; Tarr & Bülthoff, 1998). Thus, if people identify objects primarily based on unique external components, then computing the distinctiveness of these based on a few words of language is unlikely to predict age of acquisition. One way to test this explanation is to collect feature norms on novel object categories, for which names are learned in a second experiment.

As noted above, the majority of our results are based on adult retrospective age of acquisition ratings. One possible explanation for some of our results may be that when asked to judge the age of acquisition of a word, adults use feature distinctiveness as a cue to rate the word. However, the fact that the MCDI ratings also show a correlation with distinctiveness suggests that our findings are not an artifact of adults using distinctiveness as an age of acquisition rating cue. Nevertheless, our ratings are correlative in nature, meaning they show a relationship between distinctiveness and age of acquisition, but do not imply causality in one direction or another.

Our results may also be interpreted as a general effect of distinctiveness on age of acquisition, most apparent through some feature categories (e.g., external form and surface). If feature distinctiveness of an object is related to the process of word learning, then it is crucial to control for distinctiveness when designing experimental stimuli. For example, when experimentally investigating whether shape is more "important" a cue than another feature category (such as color), it is important to establish that the stimuli are similarly distinctive in terms of shape as in terms of color. It may be the case that it is easier to design stimuli that stand out in terms of shape (some of the shapes used in early shape bias studies are quite unusual), while objects may not stand out in terms of color. Presenting this stimuli to a child may result in a preference for "shape-driven learning," as that is the category that is most distinctive in context. In this sense, the inconsistencies in the shape bias literature may be influenced by stimuli not being evenly distinctive across feature categories. However, it may also be that shape is a feature category that naturally allows for distinctiveness, because it has more dimensions along which it can vary, in ways that other feature categories may not. A literature review comparing the per-category distinctiveness of stimuli used in previous shape bias studies might offer an interesting view into this problem.

It is important to note that our investigation suggests that features predict a fairly small amount of the variance in age of acquisition. Features should not be viewed as the single driving force behind acquiring early words, but a contributor working alongside a range of other, well-established correlates of age of acquisition—some of which are imageability (Gilhooly & Logie, 1980), frequency (Carroll & White, 1973), phoneme and letter length (Whaley, 1978), and language structure (Hills et al., 2010; Hills, 2012). Taken together, these suggest word learning depends not only on the properties of words and how those words are used, but also on the features of the objects to which the words refer.

Notes

1. Whether this feature similarity extends to the conceptual understanding of features (and categories) is the topic of an ongoing debate (for a review of the debate, see Elman, 2008). We do not attempt to resolve this debate, but nonetheless aim to provide a novel investigation into the relationship between feature biases, mutual exclusivity, and age of acquisition.

2. The McRae feature norms are generated by adults. They do not offer direct insight into how infants view the world, but they should be viewed as a plausible proxy representation of an infant's view. Our results, then, rely on the assumption that the adult view of an object's features is to an extent representative of an infant's.

3. Spearman's rank correlation was used due to a possible violation of the normality assumptions. A Shapiro–Wilk test showed that both the Kuperman age of acquisition data ($W = 0.97$, $p < .001$) and the MCDI age of acquisition data ($W = 0.97$, $p < .05$) are significantly different from a normal distribution.

4. Data and code to reproduce the results found in this work can be found online: https://github.com/tomasengelthaler/FeatureDistinctiveness

5. All the correlation results mentioned in this paper are Spearman's rank correlations. This is both due to the fact that the age of acquisition data is likely not normally distributed (see note 3) and also due to an absolute difference between the two types of age of acquisition (parent-indexed and adult rated, as mentioned in the method's section). When using adult-rated age of acquisition, the Spearman's rank correlation should produce results more relevant to child learning.

6. Limiting the analysis to 109 words in the MCDI in a regression predicting distinctiveness, neither MCDI (parent-indexed age of acquisition) nor the Kuperman norms (adult retrospective age of acquisition) explain additional variance after controlling for the other. This suggests that the two norms account for a common variance in explaining feature distinctiveness.

7. When looking at parent-reported age of acquisition (MCDI), no significant effects were found. The main reason for using adult-rated age of acquisition is the fact that they are much larger in size, allowing us to cover a broader range of words (gaining more statistical power). The MCDI only contains 113 words for which there are McRae feature norms. In comparison, the Kuperman adult age of acquisition norms contain 492 words overlapping with the McRae feature norms. When looking at subsets of feature categories, this difference in sample size becomes even more apparent. When interpreting our results and in comparing it with earlier work (Hills et al., 2009a), it is important to keep this difference in mind.

8. Corrections for Type I errors in multiple correlations are not straightforward, especially in cases where the correlations are unlikely to be independent. Several of our analyses have fairly low sample sizes, meaning that adjusting for multiple tests could also result in Type II errors (rejecting a significant result when there in fact

is one). Nevertheless, where appropriate, we include both unadjusted and Bonferroni adjusted p values. We encourage the reader to compare these two and make their own judgment on how robust our results are.

References

Arbesman, S., Strogatz, S. H., & Vitevitch, M. S. (2010). The structure of phonological networks across multiple languages. *International Journal of Bifurcation and Chaos*, *20*, 679–685.

Beckage, N., Smith, L., & Hills, T. (2011). Small worlds and semantic network growth in typical and late talkers. *PLoS ONE*, *6*, 1–6.

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*, *94*, 115.

Bilson, S., Yoshida, H., Tran, C. D., Woods, E. A., & Hills, T. T. (2015). Semantic facilitation in bilingual first language acquisition. *Cognition*, *140*, 122–134.

Booth, A. E., & Waxman, S. (2002). Object names and object functions serve as cues to categories for infants. *Developmental Psychology*, *38*, 948.

Booth, A. E., & Waxman, S. R. (2008). Taking stock as theories of word learning take shape. *Developmental Science*, *11*(2), 185–194.

Booth, A. E., Waxman, S. R., & Huang, Y. (2005). Conceptual information permeates word learning in infancy. *Developmental Psychology*, *41*, 491–505.

Carroll, J. B., & White, M. N. (1973). Age-of-acquisition norms for 220 picturable nouns. *Journal of Verbal Learning and Verbal Behavior*, *12*, 563–576.

Cimpian, A., & Markman, E. M. (2005). The absence of a shape bias in children's word learning. *Developmental Psychology*, *41*(6), 1003.

Collisson, B. A., Grela, B., Spaulding, T., Rueckl, J. G., & Magnuson, J. S. (2015). Individual differences in the shape bias in preschool children with specific language impairment and typical language development: Theoretical and clinical implications. *Developmental Science*, *18*(3), 373–388.

Colunga, E., & Smith, L. B. (2008). Knowledge embedded in process: The self-organization of skilled noun learning. *Developmental Science*, *11*, 195–203.

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.

Dale, P. S., & Fenson, L. (1996). Lexical development norms for young children. *Behavior Research Methods, Instruments, & Computers*, *28*, 125–127.

Dhillon, I. S., & Modha, D. S. (2001). Concept decompositions for large sparse text data using clustering. *Machine Learning*, *42*, 143–175.

Diesendruck, G., & Bloom, P. (2003). How specific is the shape bias? *Child Development*, *74*(1), 168–178.

Elman, J. L. (2008). The shape bias: An important piece in a bigger puzzle. *Developmental Science*, *11*(2), 219–222.

Gentner, D., & Imai, M. (1995). A further examination of the shape bias in early word learning. In E. V. Clark (Ed.), *Proceedings of the twenty-sixth annual child language research forum* (pp. 167–176). Stanford, CA: CSLI Publications.

Gilhooly, K. J., & Logie, R. H. (1980). Age-of-acquisition, imagery, concreteness, familiarity, and ambiguity measures for 1,944 words. *Behavior Research Methods & Instrumentation*, *12*, 395–427.

Graham, S. A., & Diesendruck, G. (2010). Fifteen-month-old infants attend to shape over other perceptual properties in an induction task. *Cognitive Development*, *25*, 111–123.

Hills, T. (2013). The company that words keep: Comparing the statistical structure of child-versus adult-directed language. *Journal of Child Language*, *40*, 586–604.

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

Hills, T., Maouene, M., Maouene, J., Sheya, A., & Smith, L. (2009b). Categorical structure among shared features in networks of early-learned nouns. *Cognition*, *112*, 381–396.

Hills, T., Maouene, J., Riordan, B., & Smith, L. (2010). The associative structure of language and contextual diversity in early language acquisition. *Journal of Memory and Language*, *63*, 259–273.

Hills, T. T., Jones, M. N., & Todd, P. M. (2012). Optimal foraging in semantic memory. *Psychological review*, *119*(2), 431.

Iversen, J. R., Patel, A. D., & Ohgushi, K. (2008). Perception of rhythmic grouping depends on auditory experiencea. *The Journal of the Acoustical Society of America*, *124*(4), 2263–2271.

Jaswal, V. K., & Hansen, M. B. (2006). Learning words: Children disregard some pragmatic information that conflicts with mutual exclusivity. *Developmental Science*, *9*, 158–165.

Jones, S. S., Smith, L. B., & Landau, B. (1991). Object properties and knowledge in early lexical learning. *Child development*, *62*(3), 499–516.

Kachergis, G., Yu, C., & Shiffrin, R. M. (2012). An associative model of adaptive inference for learning word–referent mappings. *Psychonomic Bulletin & Review*, *19*(2), 317–324.

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

Kemp, C., Perfors, A., & Tenenbaum, J. B. (2007). Learning overhypotheses with hierarchical Bayesian models. *Developmental Science*, *10*(3), 307–321.

Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods*, *44*, 978–990.

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

Logothetis, N. K., & Sheinberg, D. L. (1996). Visual object recognition. *Annual Review of Neuroscience*, *19*, 577–621.

Love, B. C., Medin, D. L., & Gureckis, T. M. (2004). SUSTAIN: A network model of category learning. *Psychological Review*, *111*, 309–332.

Markman, E. M. (1990). Constraints children place on word meanings. *Cognitive Science*, *14*, 57–77.

Markman, E. M., & Wachtel, G. F. (1988). Children's use of mutual exclusivity to constrain the meanings of words. *Cognitive Psychology*, *20*, 121–157.

Markman, E. M., Wasow, J. L., & Hansen, M. B. (2003). Use of the mutual exclusivity assumption by young word learners. *Cognitive Psychology*, *47*, 241–275.

Markson, L., Diesendruck, G., & Bloom, P. (2008). The shape of thought. *Developmental Science*, *11*(2), 204–208.

Mather, E., & Plunkett, K. (2009). Learning words over time: The role of stimulus repetition in mutual exclusivity. *Infancy*, *14*, 60–76.

Mather, E., & Plunkett, K. (2012). The role of novelty in early word learning. *Cognitive Science*, *36*, 1157–1177.

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.

Morgan, J. L. (1996). A rhythmic bias in preverbal speech segmentation. *Journal of Memory and Language*, *35*(5), 666–688.

Plunkett, K., Hu, J. F., & Cohen, L. B. (2008). Labels can override perceptual categories in early infancy. *Cognition*, *106*(2), 665–681.

Samuelson, L. K., & Smith, L. B. (1999). Early noun vocabularies: Do ontology, category structure and syntax correspond? *Cognition*, *73*, 1–33.

Samuelson, L. K., & Horst, J. S. (2007). Dynamic noun generalization: moment-to-moment interactions shape children's naming biases. *Infancy*, *11*(1), 97–110.

Schyns, P. G., Goldstone, R. L., & Thibaut, J. P. (1998). The development of features in object concepts. *Behavioral and Brain Sciences*, *21*, 1–17.

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

Sloutsky, V. M. (2010). From perceptual categories to concepts: What develops? *Cognitive Science*, *34*, 1244–1286.

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.

Sloutsky, V. M., & Fisher, A. V. (2011). The development of categorization. In B. H. Ross (Ed.), *The psychology of learning and motivation* (Vol. 54, pp. 141–166). San Diego, CA: Academic Press.

Smith, L. B., Jones, S. S., Landau, B., Gershkoff-Stowe, L., & Samuelson, L. (2002). Object name learning provides on-the-job training for attention. *Psychological Science*, *13*, 13–19.

Smith, L. B., & Samuelson, L. (2006). An attentional learning account of the shape bias: Reply to Cimpian and Markman (2005) and Booth, Waxman, and Huang (2005). *Developmental Psychology*, *42*, 1339–1343.

Soja, N. N. (1992). Inferences about the meanings of nouns: The relationship between perception and syntax. *Cognitive development*, *7*(1), 29–45.

Soja, N. N., Carey, S., & Spelke, E. S. (1991). Ontological categories guide young children's inductions of word meaning: Object terms and substance terms. *Cognition*, *38*(2), 179–211.

Tarr, M. J., & Bülthoff, H. H. (1998). Image-based object recognition in man, monkey and machine. *Cognition*, *67*, 1–20.

Tek, S., Jaffery, G., Swensen, L., Fein, D., & Naigles, L. R. (2012). The shape bias is affected by differing similarity among objects. *Cognitive Development*, *27*, 28–38.

Thiessen, E. D., & Saffran, J. R. (2007). Learning to learn: Infants' acquisition of stress-based strategies for word segmentation. *Language Learning and Development*, *3*(1), 73–100.

Weitnauer, E., Carvalho, P. F., Goldstone, R. L., & Ritter, H. (2014). Similarity-based ordering of instances for efficient concept learning. In P. Bello, M. Guarini, M. McShane, & B. Scassellati (Eds.), *36th Annual Conference of the Cognitive Science Society* (pp. 1760–1765). Austin, TX: Cognitive Science Society.

Whaley, C. P. (1978). Word—nonword classification time. *Journal of Verbal Learning and Verbal Behavior*, *17*, 143–154.

Wu, L. L., & Barsalou, L. W. (2009). Perceptual simulation in conceptual combination: Evidence from property generation. *Acta Psychologica*, *132*, 173–189.

Yurovsky, D., Yu, C., & Smith, L. B. (2013). Competitive processes in cross-situational word learning. *Cognitive Science*, *37*(5), 891–921.