Cognitive Psychology 97 (2017) 1–30

Contents lists available at ScienceDirect

Cognitive Psychology

journal homepage: www.elsevier.com/locate/cogpsych

An associative account of the development of word learning

Vladimir M. Sloutsky [a,*], Hyungwook Yim [a,b], Xin Yao [b], Simon Dennis [c]

[a] Department of Psychology, The Ohio State University, United States
[b] Unilever Corporation, Shanghai, China
[c] School of Psychology, University of Newcastle, Australia

ARTICLE INFO

Article history:
Accepted 2 June 2017

Keywords:
Word learning
Cognitive development
Semantic development

ABSTRACT

Word learning is a notoriously difficult induction problem because meaning is underdetermined by positive examples. How do children solve this problem? Some have argued that word learning is achieved by means of inference: young word learners rely on a number of assumptions that reduce the overall hypothesis space by favoring some meanings over others. However, these approaches have difficulty explaining how words are learned from conversations or text, without pointing or explicit instruction. In this research, we propose an associative mechanism that can account for such learning. In a series of experiments, 4-year-olds and adults were presented with sets of words that included a single nonsense word (e.g. dax). Some lists were taxonomic (i.,e., all items were members of a given category), some were associative (i.e., all items were associates of a given category, but not members), and some were mixed. Participants were asked to indicate whether the nonsense word was an animal or an artifact. Adults exhibited evidence of learning when lists consisted of either associatively or taxonomically related items. In contrast, children exhibited evidence of word learning only when lists consisted of associatively related items. These results present challenges to several extant models of word learning, and a new model based on the distinction between syntagmatic and paradigmatic associations is proposed.

1. Introduction

The ability to learn and use words is a fundamental aspect of the human cognitive system. Not only is word learning virtually universal among humans, but it also exhibits an early onset: children typically reach an average vocabulary of 200–300 words by 24 months (Fenson et al., 1994). Despite its early onset, from the logical point of view, word learning should be very difficult, as the learner has to solve a number of hard problems. First, when presented with a novel word accompanying a novel item, the young word learner has to solve a *mapping* problem – how to map the word in question onto the world. For example, as noted by Quine (1960), the word *gavagai* uttered while pointing to a rabbit has an infinitely large number of potential mappings, including the rabbit, its parts, its texture, its method of locomotion, or the speaker's attitude towards the rabbit, among many others. Second, the word learner has to solve a *generalization* problem – given that most of the time words can be extended to multiple tokens, it is necessary to determine the class of entities the word refers to. For example, when an object (say a rabbit) is named, even if the mapping problem is solved, the word *gavagai* has an large number of potential extensions: this particular rabbit, some rabbits, all rabbits, rabbits and cats, land animals, mammals,

* Corresponding author at: Department of Psychology, The Ohio State University, Columbus, OH 43210, United States.
E-mail address: sloutsky.1@osu.edu (V.M. Sloutsky).

http://dx.doi.org/10.1016/j.cogpsych.2017.06.001
0010-0285/

all animals, solid objects, all things, etc. Therefore, both mapping and generalization are massively underdetermined by input, even when a to-be-named object is present, pointed to, and explicitly named (such situations are often referred to as ostensive definitions).

The situation appears to be even more difficult and uncertain when a word is introduced only in a conversation or through reading (as we argue, most words are, see Gleitman, 1990, for related arguments and evidence), without pointing or explicit naming. This uncertainty persists even if we limit our focus (which we do in present work) to word forms that are the earliest to be acquired – count nouns. Despite this massive uncertainty, most children succeed in learning words. Therefore, it is reasonable to ask: How do they accomplish this?

The goal of the work reported here is to provide answers to this question and to consider our answers in the context of previously provided ones. Given the history and the importance of the problem, it is hardly surprising that there are multiple accounts of word learning (see Regier, 2005, for an excellent review). In what follows, we briefly review two classes of accounts, inference-based and association-based. We also discuss why some of these accounts may have difficulty explaining an important aspect of word learning -- learning words from context, including conversations or reading. We then present our proposal, a set of experiments designed to test ours as well as some of the previously proposed accounts, and a computational model instantiating this proposal.

1.1. Inference-based models

The first set of approaches considers word learning as an inference problem: the learner needs to select the most appropriate hypothesis as to what the word in question might mean (see Bloom, 2000; Xu & Tenenbaum, 2007, for reviews). The basic idea is that despite massive uncertainty (i.e., the same data set is compatible with multiple hypotheses), the problem can be solved if (a) one is biased in favor of some hypotheses over others and/or (b) hypotheses are weighted in terms of their support by data. An important commonality of inference-based models is that determining the meaning of a word is driven by a set of assumptions: these assumptions help the learner to converge on the correct meaning.

One proposal is that even very young word learners use their knowledge about the language and the world to select a correct word-world mapping (Markman, 1989; Xu & Tenenbaum, 2007). Some proponents of this idea suggest that word learners hold certain assumptions that constrain the number of possible word-world mappings. In particular, young children are believed to assume that count nouns (a) denote whole objects rather than parts and (b) refer to taxonomic kinds rather than to thematic groupings (Markman, 1990; Markman & Hutchinson, 1984). Therefore, if a novel object is labeled (say a dog is labeled by a word *dog*) a young word learner would assume that the word refers to the whole animal, and not to its parts. Furthermore, when presented with a novel dog, the learner would apply the assumption that words refer to taxonomic kinds (rather than to individual objects or to thematic groupings), thus extending the word *dog* to a novel instance. These assumptions should substantially reduce the hypothesis space, thereby facilitating word learning.

Although these assumptions could be helpful for solving the mapping problem, the idea of young word learners having these assumptions (or constraints) has several substantial limitations (e.g., see McMurray, Horst, & Samuelson, 2012, for a recent set of critiques). Where do these assumptions come from? How do these assumptions get relaxed when children learn synonyms, words for superordinate categories, or adjectives? And how do conflicts among the assumptions get resolved?

In addition, while these constraints are useful for solving the mapping problem, their ability to solve the generalization problem is less clear (see Xu & Tenenbaum, 2007, for a discussion). In particular, while the taxonomic assumption may suggest to the learner that the word *dog* refers to a category (and not to an individual or to a thematic grouping), the assumption offers little help in determining where the boundaries of this category are.

A proposal by Xu and Tenenbaum (2007) has complementary strengths – it offers a solution to the generalization problem by proposing a way of weighing and comparing hypotheses. According to this proposal, young word learners solve the generalization problem by using labeling data in conjunction with "sampling" assumptions (i.e., assumptions of how the teacher samples examples during naming) as well as assumptions about the structure of categories in the world. These assumptions determine prior probabilities of what a given word might mean. In particular, young word learners assume that entities belong to categories and that categories form taxonomic hierarchies (i.e., each item can be categorized at the subordinate, basic, and superordinate levels). When an object is labeled with a novel word, the learner has to decide among the possible hypotheses, and knowledge of the conceptual hierarchy determines the prior probability of each hypothesis under consideration.

These decisions are based on a Bayesian computation -- weighing each hypothesis by its prior and its likelihood. The prior [$P(h)$] reflects the learner's expectations about possible word meanings independently of the data. The likelihood [$P(X|h)$] reflects which data patterns are likely to be observed under each particular hypothesis. The posterior probability of each hypothesis [$P(h|X)$] reflects the learner's subsequent belief about the word meaning and is calculated by multiplying the prior by the likelihood.

As argued by Xu and Tenenbaum (2007), priors favor more distinctive and natural hypotheses (e.g., *cats* vs. *things with black spots*), whereas likelihoods favor hypotheses that most closely correspond to the data. As a result, under typical conditions, the distinctive and narrow classes should be favored over broader and/or less distinctive classes. For example, if a single Labrador is labeled by the word *dax*, a child would decide that the word *dax* refers to Labradors and, possibly, to other dogs: while the likelihoods may somewhat favoring the Labradors, priors favor the dogs (although both are distinctive classes, the latter could be construed as more distinctive as it is the basic-level category). However, if three Labradors are

all labeled with the word *dax*, the likelihood (as we explain below) will massively favor the narrower class (i.e., the Labrador), and, as a result, the child will settle for the narrower class and decide that the word refers only to Labradors. Finally, if three different dogs are labeled with the word *dax*, the likelihood that *dax* refers only to Labradors will be zero, thus leading the child to decide that the word refers to dogs. Although the model has been tested on the restricted problem of learning a single novel word from a few examples, Xu and Tenenbaum (2007) suggested that "this framework in principle extends to the more general problem of learning a whole lexicon from a large corpus of experience" (p. 251).

The Xu and Tenebaum's proposal is designed for handling the generalization problem, whereas its ability to handle the mapping problem is less clear. However, even with respect to the generalization problem, questions remain. In particular, in order for this proposal to work, children need to know the structure of the world (which is necessary for calculating priors and likelihoods as well as for formulating hypotheses) by the time they start learning words. How do children acquire the knowledge so early in the absence of language? It is also not clear how young word learners identify this putative structure in the world (for example, how do they know that dog is a basic-level category, whereas terrier is a subordinate-level?). And how do beginning word learners (who are toddlers and very young children) get around difficulties with class inclusion when estimating probabilities of hypotheses and likelihoods of data, given the difficulties with class inclusion that have been reported at substantially older ages (see Siegler & Svetina, 2006; Winer, 1980, for reviews). Class inclusion refers to a situation when a subset of items (s_1) is properly included in a larger set (S) as in *German shepherds are dogs*. The principal constraint of class inclusion is that the subset cannot be larger than the set (i.e., $s_1 \leq S$), something that even preschoolers often fail to appreciate. For example, when presented with three cats and seven dogs and asked whether there were more dogs or animals, young and even elementary school children tend to answer that there are more dogs. In the face of such errors, the idea of toddlers estimating probabilities of hypotheses pertaining to a sub-class and to a class (e.g., H_1 = *dax* is a dog and H_2 = *dax* is an animal) looks quite implausible. Although one might argue that children may have "implicit" knowledge of class inclusion, it is not clear how such implicit knowledge can be measured in a non-circular way.

Although the inference approach explains some important aspects of word learning, it does so with significant cost – by endowing the child with substantial knowledge of unknown and unexplained origin. This cost suggests that the inference account may not be the right framework to handle the word learning problem and that it is necessary to look for alternatives.

1.2. Association-based models

Given the massive uncertainty and associated difficulties of mapping and generalization, how can word learning proceed without the above-discussed assumptions in place? There are a number of proposals that consider word learning as a variant of associative and attentional learning, with constraints being emergent properties or results of learning rather than its starting point (Colunga & Smith, 2005; Elman, 2009; Mayor & Plunkett, 2010; McMurray et al., 2012; Samuelson & Smith, 1999; Yu & Smith, 2007, 2012; see also Regier, 2005).

One idea is that there are mutually reinforcing correlations among language, perceptual properties and category structure (i.e., distribution of perceptual features within and between categories) and reliance on these correlations may help solving the generalization problem (e.g., Colunga & Smith, 2005; Samuelson & Smith, 1999; Yoshida & Smith, 2003). For example, Colunga and Smith (2005) proposed a model attempting to explain how children develop the ability to rely on common shape when generalizing names for solid things and to rely on material when generalizing names for non-solid things. According to the model, there are regularities in the environment that are internalized in the course of learning. In particular, categories of solid objects are organized by shape (e.g., cups are cup-shaped and balls are ball-shaped), whereas material may differ both within and across these categories. At the same time, categories of non-solid objects are organized by material (e.g. water is made of water and clay is made of clay), whereas shape varies both within and across these categories. Learning associations between solidity and shape and non-solidity and material constrains lexical extension: children begin to generalize names for solids by shape and names for non-solids by material. Learning is also assisted by cross-situational statistics: although within a single episode, there are multiple competing referents, the situation is less ambiguous across multiple episodes (McMurray et al., 2012; Yu & Smith, 2012). Across situations the word "dog" is more likely to co-occur with a dog than with a cat. In sum, proponents of this account argue that word learning is a slow rather than an instantaneous process and it may be achieved by implicit means without requiring inference, problem solving, or hypothesis testing.

Another idea (e.g., Regier, 2005) is that word learning is a variant of attentional learning – as children acquire words, they learn that some aspects of the words (e.g., speaker's pitch in English) do not predict meaning, whereas other aspects (e.g., voicing of the initial consonant) do. As a result, children learn to allocate attention to those dimensions that are relevant for determining the meaning (and not to the irrelevant ones), which in turn speeds up subsequent learning.

Mayor and Plunkett (2010) also offered a solution to the mapping and generalization problem. The model (implemented as auditory and visual Self-Organizing Maps -- SOMs) attempts to provide a solution to the mapping and the generalization problems using the association formed between the auditory and the visual SOMs. The model is based on two central assumptions. First, by the time children start acquiring words, they typically master the constituent processing – parsing words out of fluent speech and forming visual categories (i.e., both abilities are pre-lexical). And second, they have the ability to participate in joint-attention activities, and this ability is also pre-lexical.

Due to the topographical organization of SOMs early in development, many neighboring units on the visual map will become active and associated with the target unit on the auditory SOM (i.e., the word). Because these neighboring units respond to objects that are similar to the target-object, graded, similarity-based generalization can be achieved. Importantly,

the model allows the development of word learning because the quality and efficiency of generalization of word–object associations are directly related to the quality of pre-lexical representations of a to-be-labeled visual category, and the quality of these representations may change in the course of development.

The third idea (McMurray et al., 2012) is that word learning includes two component processes operating on different time scales: one is a slow associative process of mapping (and subsequent generalization) and another is a fast process of competitive referent selection that transpires in a variety of "fast mapping" tasks. The former process is critical for associating visual and auditory inputs, whereas the latter allows competition to shape the strength of these connections.

The majority of the association-based models subscribe to the view that although there are constraints on word learning, these constraints are not *a priori*, but are themselves products of word learning (e.g., Smith, Jones, Landau, Gershkoff-Stowe, & Samuelson, 2002). These constraints develop in the course of initial word learning and then begin constraining subsequent word learning. For example, early noun learning results in shape bias, a tendency of young word learners to generalize nouns to items with the same or similar shape. In turn, shape bias (which is considered within this framework to be a result of learning) in conjunction with cross-situational statistics helps to solve the mapping problem (Yu & Smith, 2007). Therefore, learning is an iterative process, with early learning resulting in biases that constrain future learning. Although, as argued by Xu and Tenenbaum (2007), association-based models may have a hard time explaining how people learn overlapping word meanings (e.g., the fact that the word "animal" refers to animals, dogs, and terriers), these models do not assume that early word learning requires a pre-linguistic baby to have the structure of categories in place. Additionally, by attempting to explain the development of constraints on word learning, associative accounts make substantial progress in explaining the development of word learning.

1.3. Common limitations of inference-based and association-based models

Note that much of the theoretical work reviewed above focused on explaining word learning by ostension – with someone (presumably knowledgeable) pointing to an object and labeling this object (e.g., "this is a dax"). Even in the case of cross-situational statistical learning (when the word is introduced in the presence of multiple candidate referents), there is an explicit naming event, with candidate visual referents being present. Therefore, with rare exceptions (e.g., Elman, 1990; McMurray et al., 2012; Siskind, 1996), the majority of inference-based and associative proposals are not equipped with the machinery for handling word learning without the visual referent being present. Moreover, even in most of these exceptions (such as McMurray et al., 2012; Siskind, 1996) some corresponding event in the world is required.

However, learning through explicit naming of a visually presented referent has its challenges (Robinson & Sloutsky, 2007, 2010). In addition, such learning can account for at best a small proportion of word learning: Words are split between concrete and abstract, and abstract words represent a large portion of the lexicon (see the MRC database; Wilson, 1988). In contrast to concrete words, abstract words have no concrete external referent that could be pointed to and labeled. Furthermore, many of the concrete words fall outside the direct experience of a given child simply because they cannot be observed directly (e.g., air) or are not immediately present (e.g., a cousin who lives in another city). Indeed, one of the defining characteristics of language (as separate from communication) is its capacity to describe things that are not present either because they exist in a different time or a different place (Hockett, 1960) and humans make extensive use of this facility.

Furthermore, the majority of words are never heard in isolation, but rather appear embedded in phrasal or sentential strings (see Gleitman, 1990, for a discussion). Therefore, while explicit naming of a visually present referent may be helpful to get word learning off the ground very early in development, it does not seem to be a plausible mechanism for the learning of the substantive majority of words. Instead, it seems likely that the majority of words are learned from linguistic context. The very phenomenon of vocabulary explosion – a rapid growth in vocabulary without parents explicitly teaching children new words – points to learning from conversations or reading (e.g., Goodman, McDonough, & Brown, 1998). For example, in a sentence "We will go to the zoo and you will see a little furry *dax*. He eats bugs", the syntactic structure (i.e., word order in English) suggests that *dax* is a thing (as opposed to an action or a property), whereas the semantic context (i.e., words *zoo*, *furry* and *eat*) suggest that *dax* is an animate thing.

The proposal that the sentence context can support word learning seems particularly plausible, given that sentence context can help disambiguating the meaning of even well-known words (e.g., Elman, 2005). For example, in sentences "The shopper saved…" vs. "The lifeguard saved…" the sentence context will result in different interpretations of the word *saved*. In the next section, we consider how words could be learned from context.

1.4. Learning words from context: syntactic, semantic, and social cues to meaning

Between birth and adulthood, a typical English-speaking child learns on average about 8–10 words per day: Learning new words starts slowly, but accelerates dramatically during the second year of life (Bloom, 1973). Although some of these words may be learned by ostension (i.e., explicit naming of a visually present referent or candidate referents), most are learned from context including conversations and reading (see Goodman et al., 1998; Nagy, Herman, & Anderson, 1985). How can words be learned from such contexts?

A number of ideas have been proposed to answer this question. One general proposal (that we review in this section) is that the context provides the learner with syntactic, semantic, or social cues to meaning. Another general proposal (that we review in the next section) is that the context provides the learner with a rich network of associations that may help the

learner to figure out the meaning of the new word. Note that these ideas are not mutually exclusive and different sources may strengthen each other.

1.4.1. Syntactic cues

One idea is that children use syntactic cues to disambiguate the meaning of a novel word. The idea goes back to Brown (1957) who elegantly demonstrated that when presented with a novel word, 3–4 year-olds used the syntactic frame (e.g., "this is a sib" vs. "this one is sibbing") to determine whether the words referred to an object, action, or property. These ideas generated much empirical support (see Bloom, 2000, for an extensive review). For example, Soja (1992) found that participants could use count noun/mass noun syntax to guide their learning of novel nouns. Syntactic cues have also been shown to facilitate learning of new verbs, the process known as syntactic bootstrapping (e.g., Fisher, 1996; Gleitman, 1990; Landau & Gleitman, 1985; Naigles, 1990).

However, while syntactic cues may offer some guidance in solving the mapping problem, they offer little guidance in solving the generalization problem. Syntactic cues might be useful for solving the mapping problem because these cues can help determine the part of speech the novel word belongs to and the corresponding ontological class (e.g., object or action) or help map verbs onto events. At the same time, these cues do not help distinguishing among candidate classes within a broader ontological class (e.g., dogs vs. mammals). Therefore, syntactic cues may guide a learner to the conclusion that a particular word is a count noun and it should refer to an object, but they do not help distinguishing among candidate mappings or generalizations of the count noun. For example, if an animal is called "a dax", syntactic cues (i.e., the fact that the word is a count noun) would be helpful in determining that the word refers to an object, but they would be of little help in figuring out whether the word "dax" refers to the whole animal or to a part of the animal, and, if it refers to the whole animal, whether it refers to this particular kind of animal or to all animals.

1.4.2. Semantic cues

The second ideas is that semantic cues can assist word learning. For example, Goodman et al. (1998) demonstrated that 2-year-olds could learn novel words when given sufficient semantic cues, such as familiar verbs (e.g., "Mommy feeds the ferret") suggesting possible meanings of a novel noun.

1.4.3. Social cues

The third idea is that word learning could be assisted by the "theory of mind," i.e., by solving a social problem of what the speaker is calling attention to (e.g., Akhtar & Tomasello, 2000). This social problem could be solved by focusing on the speaker's gaze direction, facial expression (Baldwin, 1991, 1993), or on some relevant aspects of a social situation (Akhtar, 2002). For example, Akhtar (2002) presented 2-year-olds with a word learning task. In one condition, participants' attention was attracted to the texture of objects ("this is a smooth one and this is a fuzzy one"), and in another condition their attention was attracted to the shape of the objects ("This is a round one and this is a square one"). Participants were then shown a triad of novel objects, and one of the novel objects was labeled ("this is a dacky one"). The remaining objects matched the labeled object either in shape or in texture. The results indicated that the context affected inferred meaning of the new adjective – participants were more likely to infer that "dacky" referred to shape in the shape-relevant context and that it referred to texture in the texture-relevant context. However, while social problem solving may offer some assistance in solving the mapping and generalization problem, its assistance is rather limited, especially when entities have multiple feature dimensions. For example, what is the set of relevant dimensions that would help learning the word *cat*?

Although each of these proposals can assist word learning, they are generally compatible with either an inferential or an associative mechanism of word learning and thus cannot distinguish between these possibilities. In the next section, we present a proposal attempting to overcome this limitation by offering an account of word learning from context. We then present a set of experiments demonstrating that (a) children indeed learn words from linguistic context, (b) what they learn changes in the course of development, and (c) the proposed associative account (described below) can capture both learning itself and developmental differences in what is learned and it succeeds where its competitors relying on inference fail.

1.5. Learning words from context: different types of associations in word learning (the current proposal)

While we agree that each of the above mentioned cues (i.e., syntactic, semantic, or social) may contribute to word learning, the current work focuses on the contribution of various networks of associations that are learned in the course of language acquisition, and, in turn, guide learning of new words. Specifically, we propose and test a mechanistic account of how associative cues guide word learning. At the heart of the proposal is the idea that the context in which words are presented provides associative cues that trigger a candidate meaning of a novel word. We consider two types of associations that can guide word learning – syntagmatic and paradigmatic (Dennis, 2005; Ervin-Tripp, 1970; Nelson, 1977).

Syntagmatic associations refer to links between words that co-occur in close temporal proximity (such as *furry* and *dog*) as in "The *furry dog*" (Brown & Berko, 1960; Nelson, 1977). These associations are often established between words belonging to different grammatical classes (e.g., an adjective and an noun as in the example above). Therefore, words associated syntagmatically tend to be related thematically.

Paradigmatic associations refer to links between words that belong to the same grammatical class (e.g., both are nouns, such as *cat* and *dog*). These words typically do not co-occur in the same sentence (McNeill, 1963); instead these words appear

in similar sentential contexts as in "The furry *cat*" and "The furry *dog*". How and why are paradigmatic associations formed? One possibility is that paradigmatic associations are formed because syntagmatic associations lead to predictions about what will occur in a certain sentential slot. The predicted word is then associated with the actual word that appears in that slot. In the previous example, a syntagmatic association between "furry" and "cat" would be formed as a consequence of hearing "The furry cat." During the exposure to the second sentence, when the learner has processed "The furry," they would automatically make a prediction that "cat" was about to appear on the basis of the previous syntagmatic association. As a consequence, 'cat" becomes active. In fact, however, the word "dog" appears at this time, and so a paradigmatic association is formed between "cat" and "dog". Importantly, words associated paradigmatically tend to be taxonomically related.

The formation of paradigmatic associations depends on the existence of syntagmatic associations and so on average it must occur after the formation of syntagmatic associations. However, this need not be a stage like progression as some paradigmatic associations that depend upon frequent utterances may well form before the syntagmatic associations that rely on exposure to rare utterances (see Luche, Durrant, Floccia, & Plunkett, 2014, for potential evidence of emerging paradigmatic associations in infancy). In general, syntagmatic associations exhibit early onset, whereas paradigmatic associations tend to emerge later in development appearing around 6-years of age (McNeill, 1963; Nelson, 1977). Under this construal, early in development (i.e., before many paradigmatic associations come online), syntagmatic associations are the primary source providing cues to the meaning of novel words. Later in development, both syntagmatic and paradigmatic associations provide cues to novel words. If this is the case, then early in development syntagmatic associates should be better cues than paradigmatic ones.

To illustrate, imagine that the learner is presented a sequence of words "..., *furry*, *dax*". When syntagmatic associations are formed, which we believe is the case for young children, older children, and adults, *furry* is associated with the word *animal*. Therefore, *animal* will be activated, and, as a result, both *animal* and *dax* will be active at the same time, which may result in interpreting *dax* as some kind of an animal. Now imagine that the learner is presented with words "..., *cat*, *dax*." When syntagmatic associations predominate (which is the case for young children), *cat* will not activate the word *animal*, whereas when paradigmatic associations are formed (which is the case for older children and adults), *cat* will activate *animal*. Therefore, only older participants will interpret *dax* as some kind of an animal.

In other words, earlier in development, when syntagmatic associations predominate, only *furry* will be associated with *animal*, whereas later in development (when paradigmatic associations are formed), both *furry* and *cat* will be associated with *animal* (*furry* via the syntagmatic route, and *cat* via the paradigmatic route). Under this construal, the probability that the word *dax* would be considered as referring to some sort of an animal, would be proportional to associative strength between the words accompanying *dax* and the word *animal*. These associations are reflected in Forward Associative Strength (FAS) or the probability that the word *animal* will be recalled in a free association task, given the word *furry* or the word *cat* (Nelson, McEvoy, & Schreiber, 2004). If people rely on syntagmatic and/or paradigmatic associations when learning words from context, then the initial meaning of a word learned from context is broad and imprecise and additional information would be needed to zero in on a precise meaning of the word. This initially broad and imprecise meaning is potentially advantageous as it may help future learning of multiple overlapping meanings of a word.

Importantly, the formation of some (i.e., second-order) associations based on other (i.e., first-order) associations may not be limited to syntagmatic and paradigmatic associations in language. For example in studies of relational memory, when item A appears in context X (e.g., an object on a particular background) and item B appears in context X, items A and B may become associated (e.g., Preston, Shrager, Dudukovic, & Gabrieli, 2004; Zeithamova & Preston, 2010).

If many words are indeed learned from context, it is surprising that the majority of models of word learning focus on learning by ostension (i.e., with explicit naming and candidate referents being present). In the current study, we attempt to fill this gap by focusing on learning words from context. The role of context has been emphasized in models such as Latent Semantic Analysis (LSA, Landauer & Dumais, 1997), where words have similar meanings to the extent they share similar context (i.e., a paragraph in which they appear). However, unlike the proposed account, LSA is not a model of learning (i.e., similarity computations are performed when a corpus is in place), and it has no obvious way of accounting for developmental changes in word learning. In contrast, the proposed account can handle both problems by exploiting the fact that the temporal structure of sentences can give rise to syntagmatic associations.

The proposed account helps to solve an important problem of learning words from context, while presenting interesting challenges to inference-based accounts. It also provides (at least partial) solutions to the mapping and generalization problems, while not having (the discussed above) costs associated with inference-based proposals.

1.6. The current study

Research presented below consists of two parts. In the first part we present experiments designed to test the proposed associative account. In all reported experiments, a new word was embedded in a list of familiar words. We used multiple lists varied in (a) forward associative strengths (FAS) with words *animal* or *machine*; (b) taxonomic coherence, and (c) thematic coherence and examined how the lists affected lexical extensions of the novel word in children and adults. For all the lists, for adults, FAS values of each word with the word animal were derived from free association norms (Edinburgh Word Association Thesaurus database, http://www.eat.rl.ac.uk), whereas for children these values were derived empirically. In all reported experiments, we also attempted to fit the behavioral data using FAS values, Xu and Tenenbaum's (2007) Bayesian model of word learning (hereafter XTB) and the TOPICS model (Griffiths, Steyvers, & Tenenbaum, 2007). While both XTB and

TOPICS are Bayesian models, there are a number of important differences. First, the former is better suited to exploit taxonomic relatedness, whereas the latter can exploit either taxonomic or thematic relatedness. In addition, TOPICS discovers latent structure from experience, whereas XTB evaluates pre-specified hypotheses.

In Experiment 1, 4-year-olds and adults were presented with a novel word embedded either in a Taxonomic list (i.e., all members of the list were animals referred to by count nouns) or in an Associative list (i.e., all members of the list were semantic associates of the word *animal*, with some words being count nouns and some being adjectives). Word learning was tested in a subsequently presented label extension task. Different models make different predictions with respect to performance on this task and participants' performance was compared to that predicted by the reviewed models. In general, if participants rely on inference, they should be more likely to learn the word in the Taxonomic condition than in the Associative condition, whereas the reverse should be the case if word learning was driven by associations.

To examine the generality of our findings, in Experiment 2, we expanded this approach to artifact words because artifacts differ ontologically from natural kinds and could be represented differently (see Gelman & Davidson, 2013, for a discussion). In Experiment 2, child and adult participants were presented with the same task, except that both Associative and Taxonomic lists consisted of artifact words, some of which were semantic associates of the word *machine*. Again word learning was tested in a subsequently presented label extension task and performance was compared to predictions of models of word learning.

In Experiment 3, we reduced the number of words in the Associative list to four words, or just to one word. If participants learned words in Experiment 1 by using a variant of Bayesian inference, they should be sensitive to the drastic reduction of evidence in Experiment 3. Furthermore, if they use the computational machinery proposed by XTB, learning should fail because none of these words was taxonomically related to the word *animal*.

The goal of Experiments 4 and 5 was to further test the models by attempting to confuse a rational word learner. The overall idea of Experiments 4–5 is that if participants base their lexical extension on Bayesian inference, then "diluting the data" by mixing the lists (i.e., mixing animal and color words in Experiment 4 and mixing animal and artifact words in Experiment 5) will reduce the likelihoods of the relevant hypothesis, thus decreasing word learning. In contrast, our associative account predicts that such mixing should not decrease word learning as long as the relative summed FAS with either the word *animal* or *machine* remains high. Note that across the experiments, word learning was predicted by using *relative* summed FAS (i.e., summed FAS to the word animal in relation to summed FAS to the word *animal and* the summed FAS to the word *machine*).

To foreshadow, there were substantial differences between children and adults: young children learned the words only in the Associative, but not in the Taxonomic conditions, whereas adults tended to learn in both conditions. In addition, FAS-based model tended to better predict learning than its Bayesian competitors.

The goal of the second part of the paper is to explain these developmental differences and to provide a better understanding of what develops and how. To achieve these goals, we developed a model based on asynchronous learning of syntagmatic and paradigmatic associations. In addition to providing an excellent fit, the proposed model also provides a mechanistic account of why children and adults differ in their forward associates. The account helps to better understand how children learn words from context and how word learning changes in the course of development.

2. Experiment 1: Animal lists

The goal of Experiment 1 was to examine the ability of various contexts to guide word learning. To achieve this goal, we presented 4-year-olds and adults with a novel word (*dax* or *fep*) embedded in a list of familiar words. One list was Associative (i.e., most words on this list were not animals themselves, but were syntagmatic associates of the word *animal*), with most of the words having substantial Forward Associative Strengths (FAS) with *animal* in both children and adults. Another list was Taxonomic (i.e., all words on the list were animals). For adults, all words on this list were paradigmatically associated with *animal*, with substantial FAS values between each word and the word *animal*. In contrast, in children there were no paradigmatic associations, as FAS values on the Taxonomic list were equal to zero. Target words, FAS of each word, and summed FAS of each list across the age groups are presented in Tables 1 and 2. Recall that for all experiments reported here, for adults, FAS

Table 1
Forward Associative Strength (FAS) with the word "Animal" in the Taxonomic list by age group in Experiment 1.

	Children	Adults
Cat	0.00	0.02
Dog	0.00	0.01
Fish	0.00	0.01
Bird	0.00	0.00
Horse	0.00	0.04
Squirrel	0.00	0.04
Cow	0.00	0.02
Rabbit	0.00	0.01
Summed associative strength	**0.00**	**0.15**

Table 2
Forward Associative Strength (FAS) with the word "Animal" in the Associative list by age group in Experiment 1.

	Children	Adults
Creature	0.09	0.26
Farm	0.20	0.19
Giraffe	0.00	0.09
Hamster	0.00	0.13
Bear	0.00	0.09
Furry	0.07	0.21
Feeding	0.09	0.02
Zoo	0.39	0.51
Summed associative strength	**0.84**	**1.5**

values of each word with the word animal were derived from free association norms (Edinburgh Word Association Thesaurus database, http://www.eat.rl.ac.uk), whereas for children these values were derived empirically (see below). After training, participants were presented with a label extension task, in which they were given an option of extending the new word to either animals or artifacts. Our associative account predicts that learning words from a context is driven by the relative summed forward associative strength (FAS) of a list to a target word. Specifically, if the relative summed FAS drives the learning of the novel word, then children should succeed (i.e., extend the novel word to animals) in the Associative condition (summed $FAS_{animal} = 0.84$; summed $FAS_{machine} = 0$), but not in the Taxonomic condition (summed $FAS_{animal} = 0$), whereas adults would succeed in the Associative condition (summed $FAS_{animal} = 1.5$; summed $FAS_{machine} = 0$) and possibly in the Taxonomic condition (summed $FAS_{animal} = 0.15$; summed $FAS_{machine} = 0$). We then compared observed performance with these predictions as well as with predictions of the XTB and those of another Bayesian model – the TOPICS model (Griffiths et al., 2007).

2.1. Method

2.1.1. Participants

In this and other experiments presented here, child participants were recruited from daycare centers located in middle class suburbs of Columbus, Ohio, and adults were Ohio State University undergraduate students participating for course credit. Thirty 4-year-old children ($M = 53.36$ months, $SD = 3.99$ months, 11 girls), and 30 adults participated in Experiment 1. The experiment included two between-subjects conditions (i.e., Associative and Taxonomic), with 15 participants in each condition.

2.1.2. Materials, design, and procedure

The experiment was a variant of a new word learning task that included a study phase and a test phase. To increase the correspondence of this task of learning words from context, word learning was never mentioned in the study phase: participants were simply asked to listen to the list and try to remember it.

2.1.2.1. Study phase. Participants were asked to remember words in one of the two between-subjects conditions: Taxonomic or Associative (see Tables 1 and 2). The experimenter read a word list to the participants twice (each time in a random order), and participants were explicitly instructed to remember the words for later recall. There were nine words on the list, eight of which were familiar words and one was a novel word (either *fep* or *dax*). Familiarity of each word established by checking its age of acquisition in the MRC Psycholinguistic Database -- http://www.psy.uwa.edu.au/mrcdatabase/uwa_mrc.htm). In the Taxonomic condition, the familiar words (i.e., *cat*, *dog*, *fish*, *bird*, *horse*, *squirrel*, *cow*, and *rabbit*) were all count nouns and all belonged to the category of animal. In the Associative condition, most of the familiar words (i.e., *zoo*, *farm*, *furry*, *creature*, *giraffe*, *hamster*, *bear*, and *feeding*) were thematically related to the word animal. The reason that several words were taxonomically related to animal (while having FAS = 0) is that the list was selected initially on the basis of FAS norms in adults, whereas children's FAS were obtained empirically (see below). In subsequent experiments, we address this issue by creating purely thematic lists. Word lists, forward associative strength (FAS) of each word with the word *animal*, and the summed associative strengths by age group and condition are presented in Tables 1 and 2.

For 4-year-olds forward associative strength (FAS) was established in a separate experiment, in which 48 4–5-year-olds were presented with a free association task (each word in Tables 1–4 were read to them, and their free associations for each were collected). For adults, FAS values were found in the Edinburgh Word Association Thesaurus database (http://www.eat.rl.ac.uk/), which also uses data derived from the free association task.

2.1.2.2. Test phase. In the test phase participants were presented with a label extension task, which consisted of four trials. On each trial, participants were shown a set of four pictures (see Fig. 1). Each picture was 3 cm by 3 cm, with two pictures depicting animals and two depicting artifacts. Participants were asked which one was a dax. On each of the four trials,

Table 3
Forward Associative Strength (FAS) with the words Machine, Tool, or Device in the Associative list by age group in Experiment 2.

	Children	Adults
Pocket	0.00	0.01
Ticket	0.00	0.05
Automatic	0.00	0.19
Coin	0.00	0.02
Coke	0.00	0.02
Washer	0.00	0.09
Computer	0.00	0.12
Printing	0.02	0.13
Summed associative strength	**0.02**	**0.53**

Table 4
Forward Associative Strength (FAS) with the words Machine, Tool, or Device in the Taxonomic list by age group in Experiment 2.

	Children	Adults
Oven	0.00	0.00
TV	0.00	0.00
Fridge	0.00	0.00
Telephone	0.00	0.00
Car	0.00	0.00
Camera	0.00	0.00
Watch	0.00	0.00
Bicycle	0.00	0.00
Summed associative strength	**0.00**	**0.00**

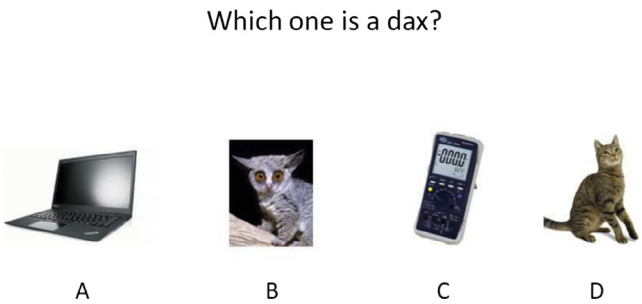

Fig. 1. An example of the Test Phase in Experiments 1–5.

participants were shown a different set of pictures. The positions of pictures in the set were pseudorandom and the order of the four trials was randomized across participants.

It was established in a separate calibration experiment with a separate group of 21 4-year-olds and 15 college undergraduates (who came from the same pool as the participants in the experiment proper) that participants had no baseline preference for either animals or artifacts. These participants were presented with a list of eight familiar color words with the word *dax* or *fep* included in the list. When presented with the described above label extension task, neither children nor adults differed from chance, both one-sample $t < 1$.

Young children were interviewed individually by a female researcher in a quiet room in their child-care centers. Undergraduate students were interviewed individually in testing rooms on campus. For all participants, the experiment was administered using Superlab 2.0 software. No feedback was given during the study or test phases.

2.2. Results and discussion

Results of the performance on the label extension task are presented in Fig. 2. One response (trial) from the taxonomic list of the children group was missing due to equipment error. A logistic regression with Age (children: 0, adult: 1) and Condition (associative: 0, taxonomic: 1) as predictors revealed a statistically significant effect of Condition (*Wald* $z = 3.01$, $p = 0.003$)

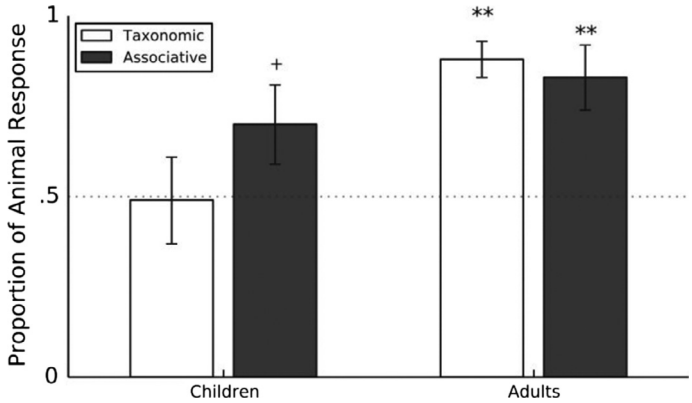

Fig. 2. Percent of animal choices in the label extension task in Experiment 1 by condition and age group. Note that +indicates $p < 0.05$, and **indicates $p < 0.001$. Error bars represent ± 1 *SEM*.

and Age × Condition interaction (*Wald* $z = 1.98$, $p = 0.048$), and a marginal effect of Age (*Wald* $z = 1.71$, $p = 0.088$). The interaction indicates that whereas adults were likely to extend the novel word to an animal in both the Associative and Taxonomic conditions (83% and 88%, respectively, both exceeding chance performance, one-sample *t*s > 3.6, one tail *p*s < 0.001, both *d*s > 0.95), young children reliably extended the novel word to animals only in the Associative condition (70%, above chance, one-sample $t = 1.8$, one tail $p = 0.045$, $d = 0.47$). In contrast to adults, their extension of novel words to animals in the Taxonomic condition (49%) did not exceed chance (one-sample $t < 1$).

In short, while adults succeeded in both conditions, young children succeeded only in the Associative condition. Failure of young children to learn words in the Taxonomic (but not in the Associative) condition is difficult to reconcile with inference accounts of word learning.

2.2.1. Comparison of predictions from word learning models

In this and other experiments, we also compared the observed results with predictions made by FAS, XTB, and TOPICS (Griffiths et al., 2007). In what follows, we explain how predictions of the models were derived. Note that there is no obvious way to perform a formal model selection and comparison here because FAS produces only a single-point estimation, which makes it impossible to calculate likelihoods. Also, as we explain below, we do not optimize any of the models, which makes RMSE scores rather uninformative. We therefore opted for a comparison of the overall pattern generated by each model.

2.2.1.1. XTB model. Predictions of the XTB account were derived on the basis of the ideas presented in Xu and Tenenbaum (2007). Recall that hypotheses of possible word meaning are weighed in accordance with the Bayes rule:

$$P(H_i|D) = \frac{P(H_i) \cdot P(D|H_i)}{P(D)} \tag{1}$$

where $P(H_i|D)$ is the posterior probability of each hypothesis (i.e., the probability of the hypothesis given data), $P(H_i)$ is the prior probability of the hypothesis, $P(D|H_i)$ is the likelihood of the data given the hypothesis, and $P(D)$ is the probability of the data (which is effectively a normalizing constant that can be disregarded for the purposes of assessing the relative probabilities of the hypotheses). Xu and Tenenbaum specify how to estimate parameters of the model. In particular, $P(H_i)$ are higher for the more distinct extensions than for the less distinct ones, whereas $P(D|H_i)$ is higher for smaller extensions than for larger ones (i.e., the size principle). According to these ideas, the hypothesis that a word accompanying a terrier with white spots refers to dogs (a distinctive set) has a higher prior probability than the hypothesis that the word refers to cats and dogs with white spots. Similarly, according to the size principle, the likelihood that the word refers to terriers is higher than the likelihood that the word refers to all dogs, all animals, or all things.

Xu and Tenenbaum (2007) derived estimates for priors and likelihoods by presenting participants with a set of animals and artifacts and asking participants to make similarity judgments. On the basis of participants' similarity judgments, a hierarchical cluster tree was created (see Xu & Tenenbaum, 2007, Fig. 7). The tree consists of multiple nodes, and as pointed by the authors, the height of each node represents the average pairwise dissimilarity of the objects in the corresponding cluster. The length of the branch above each node indicates the distinctiveness of the cluster – the average within-cluster similarity compared to similarity of the cluster's members to objects in the next nearest cluster. The tree allows an estimation of both priors and likelihood which are presented in Eq. (2) and (3) respectively.

$$P(H) = height(parent[h]) - height(h) \tag{2}$$

$$P(D|H) \propto \left[\frac{1}{height(h) + \epsilon}\right]^n \tag{3}$$

where *height*(h) is the smallest node that is compatible with the data, n is the number of episodes supporting the hypothesis (in our case, it is the number of words on the list), and ε is a small error term.

These estimates of priors and likelihoods were used to derive predictions of the Bayesian account for results of the reported experiments where an ε value of 0.05 was used as in Xu and Tenenbaum (2007). In Experiment 1, participants were presented with a forced choice between the new word referring to animals or to artifacts. Therefore, we calculated posteriors for these two hypotheses.

As in XTB, we used cluster analysis to calculate the priors and likelihoods (the details and results of our cluster analysis are presented in Appendix A). To be consistent with XTB, we applied the results of adults' clustering to children. Note that the behavior of the model could be different (but not necessarily more accurate) if we used results of children's clustering with children (cf. Jenkins, Samuelson, Smith, & Spencer, 2015).

When lists included only animals or only artifacts, the calculation of likelihoods was straightforward. When lists were mixed (e.g., Bear, Hamster, Farm, Feeding...), we had two choices. First, we could find the highest node that included all these words, in which case, the likelihood of the meaning restricted only to animals would be equal to 0. Alternatively, we could calculate the likelihood of animal by counting the number of animal words and the likelihood of artifact by counting the number of artifact words. We opted for the first option following the original proposal by Xu and Tenenbaum[1]. Once we calculated the posterior probabilities, we used a variant of the Luce choice rule (as in Eq. (4)), with a small constant ε (i.e., 1e−16) being added to each probability in order to avoid an indeterminate form (i.e., 0/0). Note that the model fit could be optimized by adding a free parameter (e.g., the "temperature" parameter) to the Luce choice rule. However, we decided against optimizing the fit for all models because our attempts to optimize resulted in an unrealistically poor performance of XTB.

$$P(Animal) = \frac{P(Animal|D) + \varepsilon}{P(Animal|D) + P(Artifact|D) + 2\varepsilon} \tag{4}$$

Unlike the other two models (i.e., FAS-based, and TOPICS), XTB does not assume a developmental difference in the model except perhaps for a stronger basic-level bias in adults. Because the experiments presented here do not contrast extension to different levels, XTB makes identical predictions for adults and children.

2.2.1.2. TOPICS model. The TOPICS model assumes that a document (i.e. a set of words) is a mixture of latent topics, with each topic generating words. The model attempts to account for a corpus of text to which it is exposed by inferring parameter distributions that make the text probable. So, the model is looking for sets of words that are interchangeable. If there is any developmental difference in how words could be learned from context, it has to come from changes in the number of topics or from the length of the unit that was considered a document (we tested entire conversations and individual utterances as possible document units)

To derive predictions of the TOPICS model, we used the CHILDES (MacWhinney, 2000) database that included both adults and children speech samples. The experimental list was added to the corpus as a document and the model was fit to each corpus using the Markov Chain Monte Carlo algorithm outlined by Stevers and Griffiths (2002). The model was then used to derive the probability of each word in the corpus given the study "document". All words in the corpus were categorized as animals, artifacts or neither by an independent observer. The probability that the participant would choose an animal in the label extension task was set to the sum of the probabilities of the animal words divided by the sum of the probabilities of animal and artifact words as in Eq. (4). In addition, we varied the document size (i.e., utterance or conversation), and the size of the topic (i.e., 50, 100, 300) resulting in six different variations of the model. We searched the best fitting models for the children data and adult data separately based on Pearson correlation coefficient value. The best fitting model was when the document size was at the conversation level with 100 topics for both adults and children (see Fig. B1 in Appendix B for correlation coefficients for all variations of the model). The document size and number of topics could be a key parameter in explaining the developmental difference in the data for the TOPICS model. However, the best fitting models for the two age groups had the same document size (i.e., conversation), and number of topics (i.e., 100). Therefore, prediction from TOPICS was identical for adults and children as XTB.

2.2.1.3. FAS-based model. We also derived predictions based on the FAS of the list. To do so, we calculated the summed forward associative strength for animal and for artifact. We then calculated the relative strength of Animal FAS using a variant of the Luce choice rule where a small constant ε (i.e., 1e−16) was added to each term in order to avoid an indeterminate form (i.e., 0/0). as in Eq. (5).

$$P(Animal) = \frac{FAS_{Animal} + \varepsilon}{FAS_{Animal} + FAS_{Artifact} + 2\varepsilon} \tag{5}$$

Predictions of the XTB, TOPICS, and FAS, and the observed data are presented in Fig. 3a for the Taxonomic condition and 3b for the Associative condition (see also Figs. 10–12, for the overall fit of each model across all the experiments). Here and in other experiments we provide qualitative description of the ability of the models to fit the data. After presenting individual

[1] Note that the second option somewhat improved model predictions (i.e., correlation coefficients increased from 0.313 to 0.369). However, these values, while requiring additional assumptions, are still the worst among the three compared models.

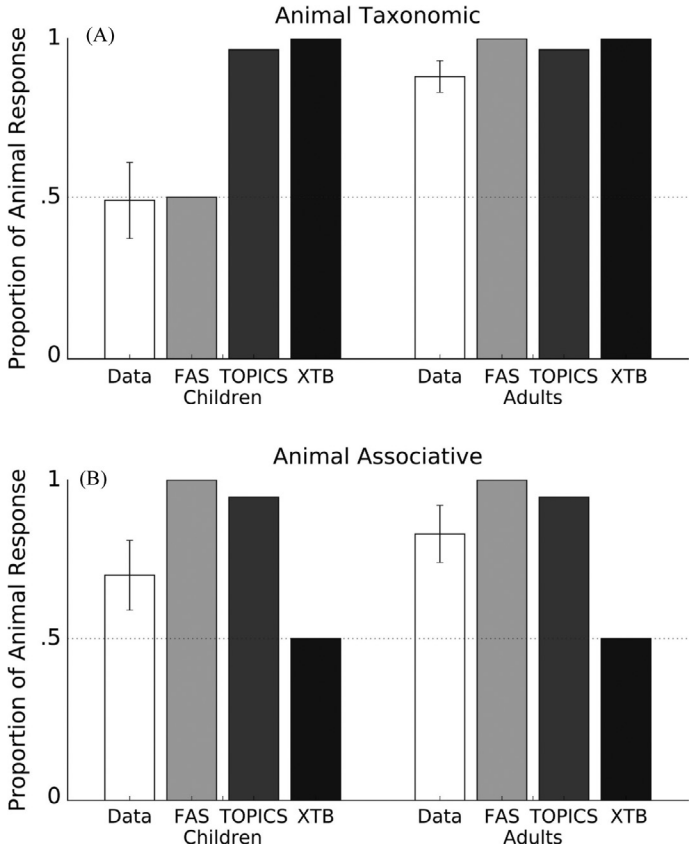

Fig. 3. Predictions of models and observed data in Experiment 1. (A) Animal taxonomic list. (B) Animal associative list. Error bars represent ± 1 *SEM*.

experiments (predictions for Experiment 1 are presented in Fig. 3), we will discuss the ability of each model to fit the entire data set.

In the Taxonomic condition (see Fig. 3a), all the models predicted adults' performance well, whereas only the FAS-based model accurately predicted children's performance (the two other models dramatically overestimated it). In the Associative condition (see Fig. 3b), XTB predicted no learning, whereas FAS and TOPICS predicted successful learning for both age groups. To test the generality of these findings, in Experiment 2 we used the same task as in Experiment 1, but presented participants with artifact lists.

3. Experiment 2: Artifact lists

3.1. Method

3.1.1. Participants

Thirty-one 4-year-old children (*M* = 55.2 months, *SD* = 3.52 months, 19 boys and 12 girls) and 30 adults participated in Experiment 2, none of whom had participated in Experiment 1. The experiment included two between-subjects conditions (i.e., Associative and Taxonomic), with 16 subjects in the Children-Associative condition, and 15 subjects each in other conditions.

3.1.2. Materials, design, and procedure

The design and procedure of Experiment 2 were identical to those in Experiment 1, with the only difference being in the set of stimuli used in Experiment 2. In contrast to Experiment 1, the eight familiar words were artifacts (familiarity of each word was established by checking its age of acquisition in the MRC Psycholinguistic Database -- http://www.psy.uwa.edu.au/mrcdatabase/uwa_mrc.htm). Similar to Experiment 1, for 4-year-olds forward associative strength (FAS) was established in a separate experiment, in which 45 4–5-year-olds were presented with a free association task. For adults, FAS values came from the Edinburgh Word Association Thesaurus database (http://www.eat.rl.ac.uk/), which also uses data derived from the free association task. FAS values represent the proportion of the words *Machine, Tool, or Device* generated in response to each word in Tables 3 and 4.

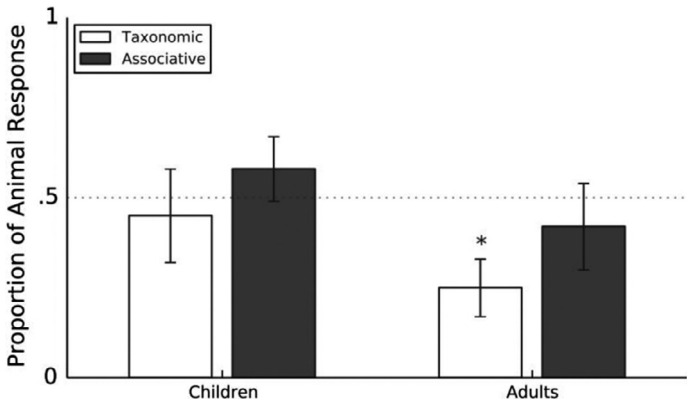

Fig. 4. Percent of animal choices in the label extension task in Experiment 2 by condition and age group. Below chance performance indicates choice of artifacts as candidate references of the novel word. Note that *indicates $p < 0.005$. Error bars represent ± 1 *SEM*.

3.2. Results and discussion

Results of performance on the label extension task are presented in Fig. 4. A logistic regression with Age (children: 0, adult: 1), and Condition (associative: 0, taxonomic: 1) as predictors showed only a marginal effect of Age (*Wald* $z = -1.79$, $p = 0.074$) with no effects for Condition nor Age × Condition interaction (*Wald zs* < −1.42, *ps* > 0.155). In particular, adults were more likely to extend the novel word to artifacts than young children. Young children exhibited chance performance across the conditions (one-sample *ts* < 0.83), whereas, adults consistently extended the novel word to artifacts at least in the Taxonomic condition (75%, exceeding chance performance, one-sample $t = 3.09$, one tail, $p = 0.004$, $d = 0.80$).

3.2.1. Comparison of predictions from word learning models

Predictions of the XTB, TOPICS, and FAS-based models, and the observed data are presented in Fig. 5a for the Taxonomic condition and 5b for the Associative condition. In the Taxonomic condition (see Fig. 5a), only FAS-based model predicted children's failure to learn, whereas XTB overpredicted performance, and TOPICS predicted animal responding. For adults, only XTB predicted successful learning, whereas FAS-based model underpredicted it, and TOPICS predicted animal responding. In the Associative condition (Fig. 5b), only XTB accurately predicted failure to learn in children and adults. Overall, results of Experiment 2 were somewhat more ambiguous than those of Experiment 1. We therefore provided further tests of the models in subsequent experiments.

One prediction of the XTB account is that longer lists provide stronger support for more narrow hypotheses. Also, according to the XTB account, shorter lists should result in broader hypotheses and thus in more diffused word learning (we discuss this issue in Experiment 3). In contrast, the FAS account predicts robust word learning as long as there is a sufficient summed forward associative strength between words(s) on a list and the target word (but not the competitor target). To examine these issues, we substantially shortened the lists in Experiment 3 compared to Experiments 1–2.

4. Experiment 3: Shortening the list

In Experiment 3, we shortened the associative list by leaving only four words (Experiment 3A) or only a single word (Experiment 3B). These were non-animal words that were the strongest associates of the word *animal*. These associates were likely to be syntagmatic because none of the words referred to an animal (or even to an object), yet all were associated with the word *animal*. This shortening and elimination of animals from the list should result in the likelihood that the novel word referred to an animal equaling zero in the XTB model. Therefore, according to XTB participants should not be extending the novel word to animals. In contrast, the high relative summed FAS suggests that participants should be extending the novel word to animals.

5. Experiment 3A

In this experiment, we shortened the associative list to only four words (i.e., *zoo*, *farm*, *furry*, and *feeding*), none of which referred to an animal. If word learning is driven by the XTB Bayesian process then extension of the new word to an animal in Experiment 3A should be low, as none of the words is an animal. In contrast, the FAS-based model predicts robust learning as the summed FAS remained high in both children and adults (see Table 5).

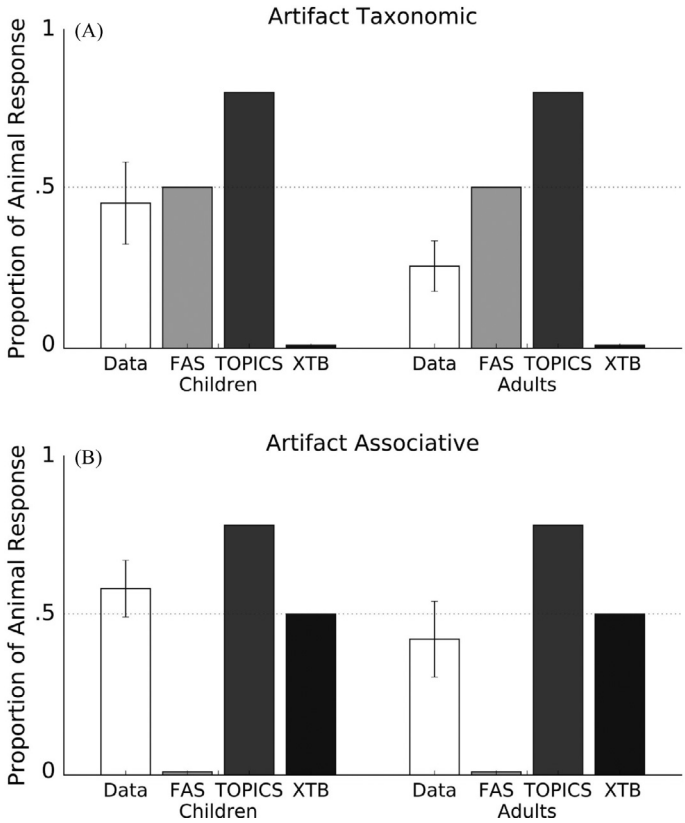

Fig. 5. Predictions of models and observed data in Experiment 2. (A) Artifact taxonomic list. (B) Artifact associative list. Error bars represent ± 1 *SEM*.

Table 5
Shortened word list used in Experiment 3A and Forward Associative Strength (FAS) with the word Animal by age group.

	Children	Adults
Zoo	0.39	0.51
Farm	0.20	0.19
Furry	0.07	0.21
Feeding	0.09	0.02
Summed associative strength	**0.75**	**0.93**

5.1. Method

5.1.1. Participants

Participants were 13 4-year-old children (*M* = 54.69 months, *SD* = 3.95 months, 10 boys and 5 girls), and 16 adults, none of whom had participated in previous experiments reported here.

5.1.2. Materials, design, and procedure

Design and procedures were identical to that of Experiment 1, with one critical difference. In Experiment 3A, the associative list was shortened to 4 associates of the word *animal*. Only the words that had strongest associations and did not denote an animal were chosen. These words were *furry*, *feeding*, *zoo*, and *farm*, with the summed associative strength of the list equal to 0.75 for children and 0.93 for adults (recall that in Experiment 1, the summed associative strength was equal to 0.84 for children and 1.5 for adults).

5.2. Results and discussion

Results are presented in Fig. 6A. A logistic regression with Age (children: 0, adult: 1) as a predictor revealed a significant effect of Age (*Wald* $z = 2.15$, $p = 0.032$), with adults exhibiting somewhat more reliable extension of the novel word to

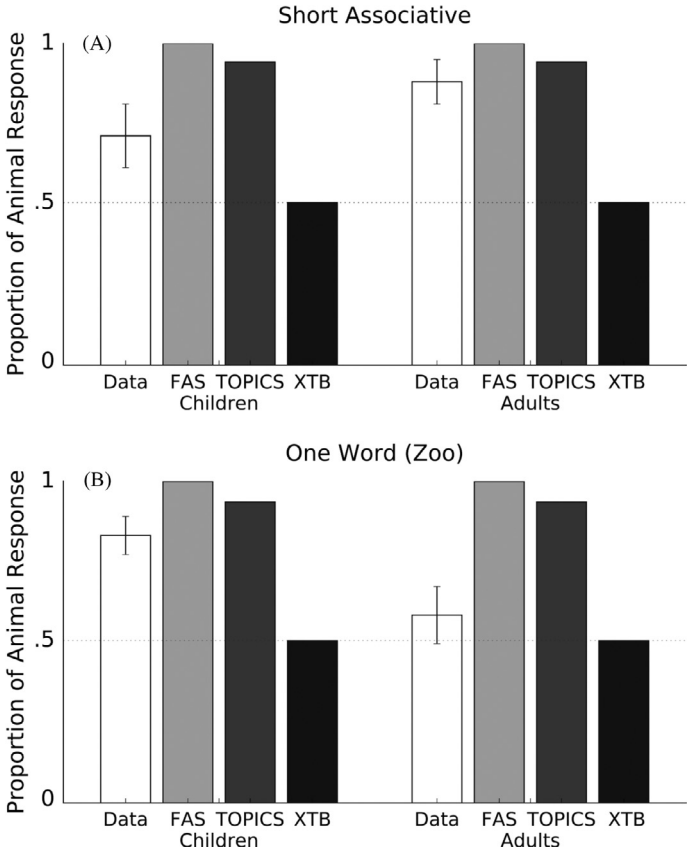

Fig. 6. Predictions of models and observed data in Experiment 3. (A) Shortened animal associative list (Experiment 3a). (B) One word animal associative list (Experiment 3b). Error bars represent ± 1 *SEM*.

animals than children (88% vs. 71%, both above chance, one-sample *ts* > 2.17, *ps* < 0.03, *ds* > 0.60). These findings replicate and further extend findings of Experiment 1. The associative list was shortened markedly (with all animal names being removed from the associative list). Whereas the likelihoods of the animal and artifact hypotheses were exceedingly low, the summed associative strength in Experiment 3A (0.75 and 0.93) was comparable to that in Experiment 1 (0.84 and 1.5). The shortening of list did not significantly affect participants' responses, with participants reliably extending the novel word to animals. As shown in Fig. 6, in the shortened list, FAS and TOPICS accurately predicted learning in children and adults, whereas XTB underpredicted learning in both groups.

To further examine the issue, in Experiment 3B, we shortened the list even more, including only one word to accompany the word dax. The word zoo was selected to retain the high associative strength between the single word on the list and the word animal. According to the proposed associative account, the proportion of animal-based generalization should remain high although it could decrease some due to the overall decrease in FAS.

6. Experiment 3B

6.1. Method

6.1.1. Participants

Participants were 15 4-year-old children (*M* = 55.2 months, *SD* = 3.47 months, 10 boys and 5 girls) and 28 adults, none of whom had participated in previous experiments reported here.

6.1.2. Materials, design, and procedure

Design and procedures were similar to those in Experiment 1 except that only the word "zoo" (FAS = 0.39 for children and FAS = 0.51 for adults) was accompanying the novel word *dax*. Therefore, the word list in this experiment only contains two words, one familiar associate of the word animal and one novel word.

6.2. Results and discussion

Proportions of animal choices are presented in Fig. 6B. Two responses (trials) each from two different adult participants were missing due to equipment error. A logistic regression with Age (children: 0, adult: 1) as a predictor revealed a statistically significant effect of Age (*Wald* $z = -3.23$, $p = 0.001$), with children exhibiting a somewhat more reliable label extension than adults. Although there was only one associative word on the list, children still exhibited reliable label extension of the word *dax* to animals (83%, one-sample $t = 5.739$, one tail $p < 0.001$, $d = 1.48$). However, adults were at chance (58%, one-sample $t = 0.908$, one tail $p > 0.19$). This low performance is surprising, and at present we do not have an explanation. One possibility is that this is a statistical aberration (which is possible, given the results of Experiment 3A). Another possibility is that adults who were exposed to the "*dax*, *zoo*" sequence might have interpreted *dax* as an adjective, whereas adults who were exposed to the "*zoo, dax*" sequence might have interpreted *dax* as a verb. Both interpretations make *dax* incompatible with choice options (i.e., objects) offered in the label extension task.

To provide further evidence distinguishing between the accounts, we conducted Experiment 4. In Experiment 4, the word *dax* was embedded in a mixed list consisting of semantic associates of the word "animal" and unrelated color words (*black*, *red*, *blue*, and *green*). This condition provides an important control by keeping FAS equivalent to that in Experiment 3A, while keeping list lengths equivalent to that in Experiment 1.

Perhaps more importantly, the color words are neither associated with the word *animal*, nor are they related taxonomically or thematically to animacy. Inference-based learning should be difficult because (a) the data suggesting that *dax* is an animal (i.e., animal words) are diluted by non-animal words and (b) it should be difficult to glean the experimenter's intent from such a mixed list. At the same time, given the high summed associative strength of the entire list, the associative approach predicts robust word learning.

7. Experiment 4: Mixed animal and color word list

The goal of Experiment 4 was to examine learning of words presented in a mixed list. In Experiment 4, the novel word was embedded in a word list consisting of eight words, four of which had high syntagmatic associative strength with the word animal (similar to Experiment 3A, the summed associative strength was equal to 0.75 for children and 0.93 for adults). Another four words were color words, which had no associative strength with the word animal. The inclusion of these color words may make rational inference less straightforward, thus leading to a decrease in animal-based generalizations. This however should not be the case, according to the associative approach.

7.1. Method

7.1.1. Participants

Participants were 15 4-year-old children ($M = 55.3$ months, $SD = 3.099$ months, 9 boys and 6 girls), and 15 adults, none of whom had participated in previous experiments reported here.

7.1.2. Materials, design, and procedure

Design and procedures were similar to that of Experiment 1 except that the list included 4 words used in Experiment 3A (i.e., *Farm*, *Zoo*, *Feeding*, and *Furry*) and 4 neutral words (*Black*, *Red*, *Blue*, and *Green*). Therefore, the summed associative strength of the list (i.e., 0.75 for children and 0.93 for adults) remained the same as that in Experiment 3A (see Table 6).

7.2. Results and discussion

The results are presented in Fig. 7. A logistic regression with Age (children: 0, adult: 1) as a predictor also shows a statistically significant effect of Age (*Wald* $z = 2.17$, $p = 0.030$): whereas both adults (92%, one-sample $t = 6.17$, one tail $p < 0.001$,

Table 6
Mixed word list used in Experiment 4 and Forward Associative Strength (FAS) with the word Animal by age group.

	Children	Adults
Zoo	0.39	0.51
Farm	0.20	0.19
Furry	0.07	0.21
Feeding	0.09	0.02
Black	0.00	0.00
Red	0.00	0.00
Blue	0.00	0.00
Green	0.00	0.00
Summed associative strength	**0.75**	**0.93**

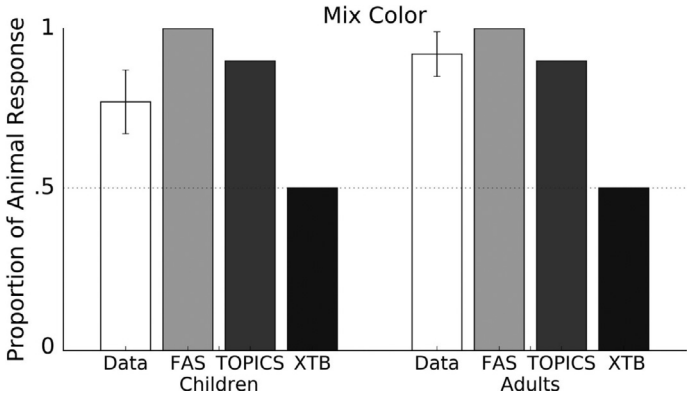

Fig. 7. Predictions of models and observed data in Experiment 4: A mixed color words and animal associative words list. Error bars represent ± 1 *SEM*.

$d = 1.59$) and children (77%, one-sample $t = 2.54$, one tail $p = 0.012$, $d = 0.66$) reliably extended the novel word to animals, adults did so more than children. While these results are compatible with the proposed associative account, they are difficult to reconcile with the XTB Bayesian account (see Fig. 7). In particular, it is not clear why participants would prefer H_2 (i.e., that *dax* is artifact) or H_1 (i.e., that *dax* is an animal) if the likelihood of either hypothesis is near zero. It is also not clear how participants could glean the intent of the experimenter to communicate that *dax* is an animal given that none of the words on the list was an animal and half of the words were not related to animals.

Simulation results show that FAS and TOPICS accurately predicted above chance performance of children and adults, whereas the XTB model failed to predict learning in both groups.

8. Experiment 5: Mixed associative list

The goal of Experiment 5 was to make a situation even less straightforward for rational inference. To achieve this goal, we mixed words from associative machine and associative animal lists.

8.1. Method

8.1.1. Participants

Participants were 15 4-year-old children ($M = 52.59$ months, $SD = 3.43$ months, 8 boys and 7 girls), and 15 adults, none of whom had participated in previous experiments reported here.

8.1.2. Materials, design, and procedure

Design and procedures were identical to that of Experiment 4 with one difference. Instead of four neutral words, four associative machine words (pocket, coin, coke, printing) were added to the four associative animal words. The four machine words were selected as they had higher associative strength with the word "machine", at least for adults. In addition, these words themselves were not in the category of machine. Therefore, the summed associative strength with the word "animal" of the list was 0.75 for children and 0.93 for adults as in Experiment 3A (see Table 7). And the associative strength with the word "machine" of the list was at 0.02 for children and 0.18 for adults.

Table 7
Mixed word list used in Experiment 5 and Forward Associative Strength (FAS) with the word Animal by age group.

	Children	Adults
Zoo	0.39	0.51
Farm	0.20	0.19
Furry	0.07	0.21
Feeding	0.09	0.02
Pocket	0.00	0.00
Coin	0.00	0.00
Coke	0.00	0.00
Printing	0.00	0.00
Summed associative strength	**0.75**	**0.93**

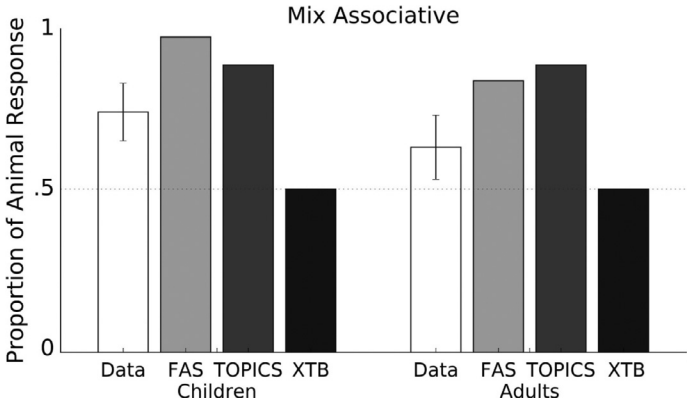

Fig. 8. Predictions of models and observed data in Experiment 5: A mixed associative animal and associative artifact words list. Error bars represent ± 1 *SEM*.

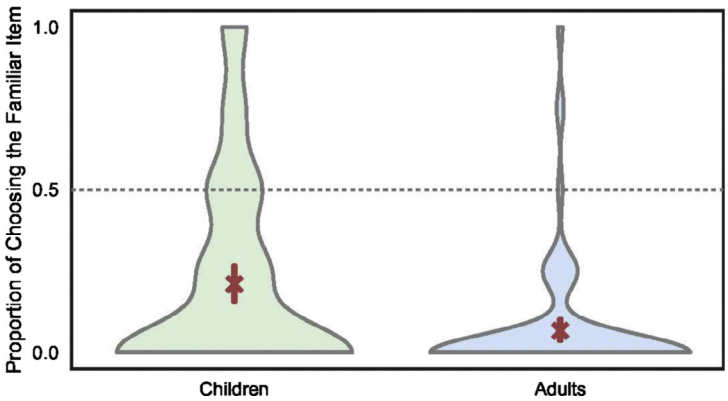

Fig. 9. Participants' proportion of choosing the familiar item at test across all conditions. The violin plot represents the distribution of the responses, the cross marks represent the mean values, and the error bars represent ± 1 *SEM*.

8.2. Results and discussion

Results are presented in Fig. 8. A logistic regression with Age (children: 0, adult: 1) as a predictor revealed no effects of Age (*Wald* $z = -1.24$, $p = 0.216$), with both children and adults extending novel words to animals. Overall, FAS and TOPICS predicted successful learning in children and adults, whereas XTB failed to do so for both age groups (see Fig. 8).

9. Analyses of participants' choices across experiments

Recall that in each experiment participants had four choices: two animals (Familiar and Novel) and two artifacts (Familiar and Novel). In previous sections, we focused on participants' choices of animals and artifacts. In this section, we focus on participants' choices of novel vs. familiar items (see Fig. 9). Importantly, both children and adults selected the novel option substantially more frequently than would be expected by chance (both *ts* > 4.63, *ps* < 0.001, Cohen's *ds* > 0.58), and, perhaps more importantly, logistic regression with Age as predictor indicated that this tendency was significantly more pronounced in adults (*Wald* $z = 6.305$, $p < 0.001$). These findings suggest that (a) both children and adults rely on mutual exclusivity (i.e., the idea that novel names correspond to novel rather than familiar objects) and (b) the tendency to rely on mutual exclusivity increases with age. The latter point is of utmost importance as the fact that mutual exclusivity increases with age (and experience) suggests that mutual exclusivity can emerge as a result of learning. Of course, having only two age groups makes this point highly speculative, and more research is needed to reach more definitive conclusions.

10. Models' predictions across experiments

Our next goal was to compare the overall fit of the models across the experiments. As explained above, there was no obvious way to perform a formal model selection and comparison here. We therefore opted for (a) a comparison of the overall

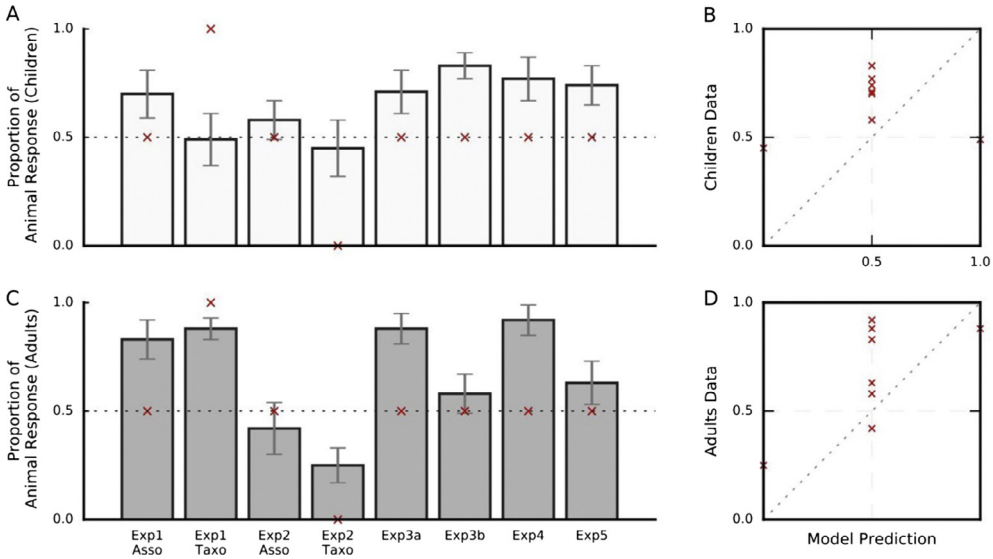

Fig. 10. Model fit of the XTB model for all 16 experimental conditions ($r = 0.449$). (A) Children's data (presented in bars) and model predictions (in red crosses), (B) Children's data (y-axis) and model prediction (x-axis) presented on a scatter plot, (C) Adults' data (presented in bars) and model predictions (in red crosses), (D) Adults' data (y-axis) and model prediction (x-axis) presented on a scatter plot. Error bars represent ± 1 *SEM*.

pattern generated by each model and (b) calculating correlations between outcomes predicted by each model and observed outcomes (see Figs. 10–12 for the overall fits of the three models). These were Pearson correlations calculated for predicted and observed mean of each condition ($N = 16$). Note that correlation coefficients on group means could be higher than those on individual values. However, here we focus on comparing the relative difference between the models' predictions and not on the absolute value of each correlation coefficient. The results indicate that FAS provided a better account of data ($r = 0.715$) than the two powerful Bayesian models – XTB ($r = 0.449$) and TOPICS ($r = 0.657$). It is worth noting that XTB and TOPICS do not make different predictions for the two age groups, while FAS had different predictions for the two age groups. In terms of overall patterns, XTB predicts learning ONLY in the Taxonomic condition, predicting no learning in the other conditions. As a result, it fails to predict successful learning in most of the associative conditions and children's failure

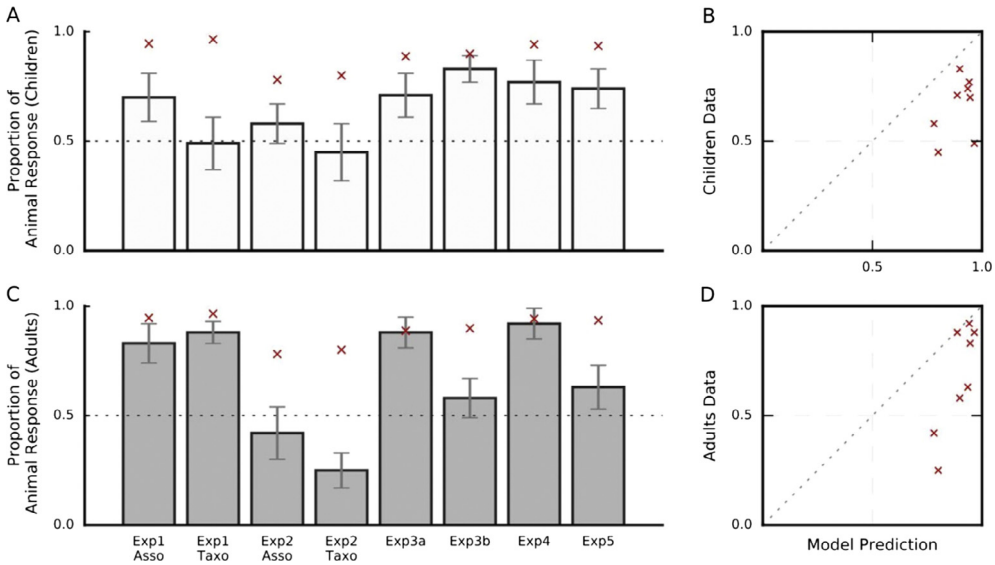

Fig. 11. Model fit of the TOPICS model for all 16 experimental conditions ($r = 0.657$) with children and adults using conversation as a document of having 100 topics. (A) Children's data (presented in bars) and model predictions (in red crosses), (B) Children's data (y-axis) and model prediction (x-axis) presented on a scatter plot, (C) Adults' data (presented in bars) and model predictions (in red crosses), (D) Adults' data (y-axis) and model prediction (x-axis) presented on a scatter plot. Error bars represent ± 1 *SEM*.

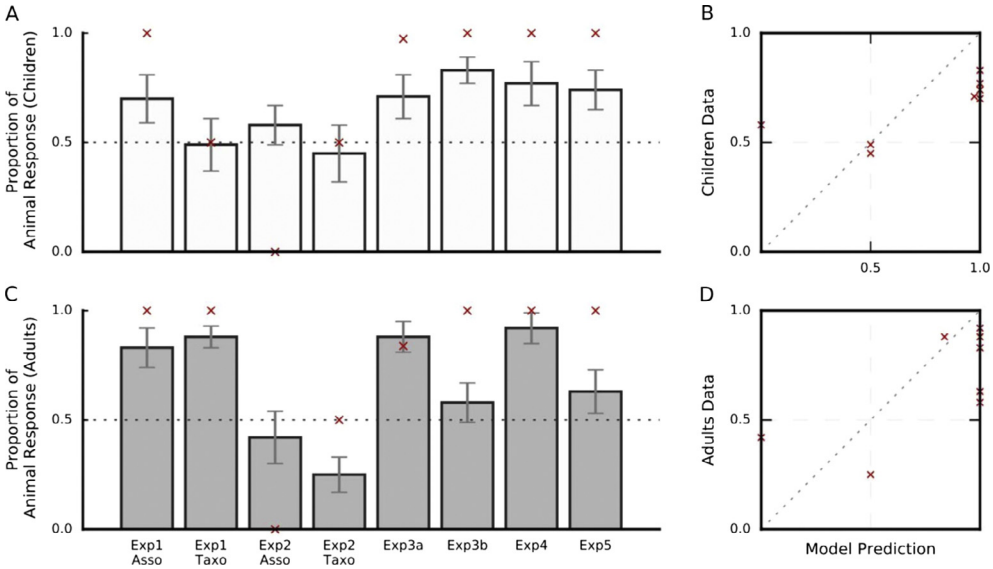

Fig. 12. Fit of the FAS-based prediction for all 16 experimental conditions ($r = 0.715$). (A) Children's data (presented in bars) and predictions (in red crosses), (B) Children's data (y-axis) and prediction (x-axis) presented on a scatter plot, (C) Adults' data (presented in bars) and predictions (in red crosses), (B) Adults' data (y-axis) and prediction (x-axis) presented on a scatter plot. Error bars represent ± 1 *SEM*.

to learn in the taxonomic conditions. TOPICS also failed to capture the overall pattern, predicting a strong animal response, regardless of the condition (including the two artifact conditions). Although FAS too failed to predict several conditions, its predictions were more accurate and failures less systematic.

Taken together these results suggest that Bayesian (and perhaps other inference) accounts may not be the right framework for explaining learning of words from context. These results demonstrate clear developmental differences (i.e., only adults, but not children learned words in the Taxonomic lists), and they suggest that summed FAS may explain the differences.

Although the FAS-based model makes reasonably accurate predictions (and these predictions differ between the age groups), it takes FAS as input, without explaining developmental differences in FAS values. The next section addresses these questions: we attempt to understand how these differences may emerge in the course of individual development. In doing so, we propose and test a model that provides a mechanistic account of the development of word learning based on the distinction between syntagmatic and paradigmatic associations. The fundamental idea of the model is experience-based development: paradigmatic associations (which capture taxonomic relationships) emerge as a result of experience with syntagmatic associations (which are first-order associations).

11. A new developmental account of word learning

The proposed model is based on a number of key ideas. First, we propose that the knowledge that children acquire as they learn words takes the form of syntagmatic and paradigmatic associations. Syntagmatic associations are formed between words that occur in close temporal proximity. An association between "zoo" and "lion" might be formed because children hear stories that involve zoos and lions. Paradigmatic associations are formed between words that occur within similar sentential contexts. An association between "tiger" and "lion" might be formed because these words fill the same slots in sentences like "The X was chasing the antelope."

While syntagmatic associations are likely to be formed between words that are thematically related, paradigmatic associations tend to form between words that are taxonomically related. If one considers the sentence "The X was chasing the antelope" there are only certain words that are likely to fill the X slot and these will be relatively large carnivores. If, however, the common context was "The X was chasing" then the class of possible paradigmatic associates expands to all animals. In this case, the taxonomic class is expanded, but it remains very likely that the filler in the X position will be taxonomically related to lion and tiger.

Before a paradigmatic association can be formed, the learner must use prior experience in similar sentential contexts to predict which words could have filled the same slot. By contrast, syntagmatic associations can be formed as soon as the two words co-occur. Consequently, we would anticipate that the impact of syntagmatic associations in word learning would be evident earlier in development than that of paradigmatic associations. It is this aspect of the model that we propose explains the inability of children to utilize taxonomic cues in the label extension task, and the subsequent ability of adults to generalize on the basis of these cues.

11.1. Model description

The model consists of two artificial neural networks - the syntagmatic network and the paradigmatic network. The syntagmatic network is trained to learn the association strength among the co-occurring words in an utterance (e.g., *zoo* and *lion*). The paradigmatic network learns the association strength between words that appeared in similar contexts, where context is defined as the word appearing before and the word appearing after the target word.

11.1.1. Training corpus

The model was trained using the CHILDES database (MacWhinney, 2000). The corpus had 1,915,095 English utterances and included speech from both children and adults. For computational efficiency, the corpus was filtered. Words that had a word frequency of at least 7 were maintained in the corpus (i.e. 12,426 unique words), low frequency words (<7) that had a frequently used suffix (see Table C in Appendix C) were converted to a general form which retained this suffix (i.e. 54 unique forms), and low frequency words that did not have this suffix information were converted to a general form that did not retain any information about the word (i.e. *xxx*). For example, the word '*silliness*' occurred fewer than 7 times, but had the noun suffix '*-iness*'. Therefore, '*silliness*' was converted to a general form '*xxxiness*' (i.e. filler (*xxx*) + suffix (*-iness*)). On the other hand, the word '*acknowledge*' occurred fewer than 7 times but did not have a frequently used suffix. Therefore, the word '*acknowledge*' was converted into '*xxx*'. By reducing the number of unique words in the corpus using this method, we were able to increase computational efficiency, while retaining some morphological information.

11.1.2. Syntagmatic network

The syntagmatic network has two layers – (1) the input layer, and (2) the output layer (see Fig. 13). The input layer has 12,426 units which is the total number of unique words in the filtered CHILDES corpus, and the output layer has 1233 units which is the total number of animal and artifact words in the corpus (i.e. 555 animal units, and 678 artifact units). The number of units in the output layer was reduced to represent only animal and artifact words for computational efficiency. Each unit represents the frequency of the word in an utterance. For example, in the utterance "the dog chased the car", the units representing 'dog', 'chased', and 'car' will have a value of 1, 'the' will have a value of 2, and other units will have a value of zero. The input layer and the output layer are fully connected and have a linear activation function. Therefore, the activation of the *j*th output unit (o_j) is

$$o_j = \sum_{i=1}^{n} (w_{ij} \cdot x_i) \tag{6}$$

where w_i is the connection weight between the *i*th input unit and *j*th output unit, x_i is the value of the *i*th input unit, and n is the total number of input units.

The network is an auto-associator, where the input activation pattern learns to predict its own pattern at the output layer. However, since the output layer only has units representing animal and artifact words, the network learns to predict animal or artifact words instead of the whole utterance pattern. For example, in the utterance "the dog chased the car", the whole input activation pattern will learn to predict the word 'dog', and 'car' (see Fig. 13).

Weights between the units are updated every utterance using a delta rule:

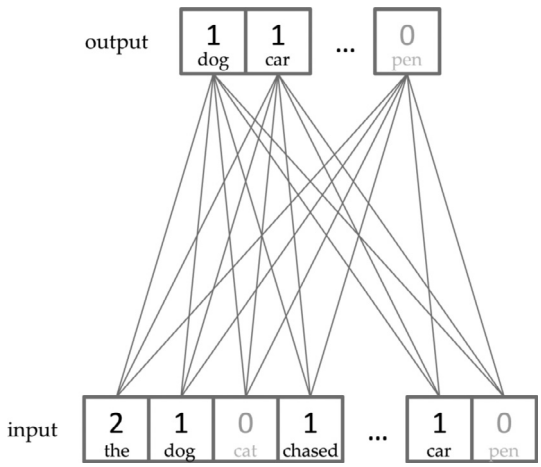

Fig. 13. The architecture of the syntagmatic network. Values depict the activation pattern of the input layer and the target pattern of the output layer when the utterance "the dog chased the car" is trained. Units that represent the words that are not in the utterance but included in the corpus such as "cat", or "pen" are set to zero.

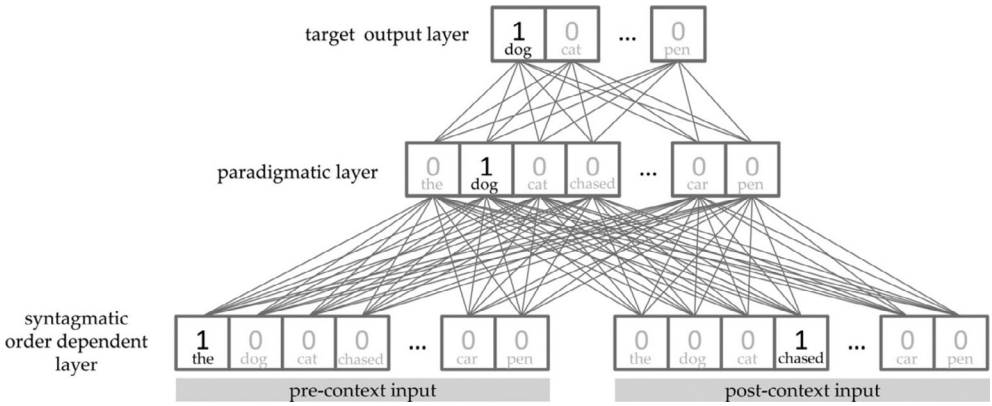

Fig. 14. The architecture of the paradigmatic network. The paradigmatic network had three layers - (1) the syntagmatic order-dependent layer, (2) the paradigmatic layer, and (3) the target output layer.

$$\Delta w_{ij} = \alpha \cdot (t_j - o_j) \cdot x_i \tag{7}$$

where Δw_{ij} is the value updated in the connection weight between the ith input unit and jth output unit, α is the learning rate which was fixed to 0.0001, t_j is the target value that the jth output unit has to learn, o_j is the value of the jth output unit, and x_i is the input value of the ith input unit. The initial values of all connection weights are set to zero.

11.1.3. Paradigmatic network

The paradigmatic network has three layers – (1) the syntagmatic order-dependent layer, (2) the paradigmatic layer, and (3) the target output layer (see Fig. 14). The syntagmatic order-dependent layer is connected to the paradigmatic layer and the paradigmatic layer is connected to the target output layer. The syntagmatic order-dependent layer has 24,852 units where the first 12,426 units represent the word before the target word, and the remaining 12,426 units represent the word after the target word. The paradigmatic layer has 12,426 units, while the target output layer has 1233 units. Each unit represents the frequency of the word in an utterance as in the syntagmatic network.

The paradigmatic associations are learned by propagating activation from the context input layer to the paradigmatic layer, and then propagating activation from the paradigmatic layer to the target output layer. A target word is defined as each word in an utterance, and the context of the target word is defined as the word occurring before (pre-context) and after (post-context) the target word. For example, in the utterance "the dog chased the car", the target word 'dog' would have the pre-context as 'the' and post-context as 'chased'. Capturing order-dependent context in this way is certainly a substantial simplification of what one might expect the real system to incorporate, and yet there have been multiple demonstrations of the power of immediate context used in similar ways (Jones & Mewhort, 2007; Redington, 1998). As we demonstrate below, even this crude approximation is able to extract a great deal of relevant information. The two parts of the context input layer independently learn to predict the target word using a delta rule (i.e. error is not backpropagated and the intermediate units are not hidden).

$$\begin{aligned} \Delta w_{ij}^{pre} &= \alpha \cdot (t_j - o_j^{pre}) \cdot x_i^{pre} \\ \Delta w_{ij}^{pos} &= \alpha \cdot (t_j - o_j^{pos}) \cdot x_i^{pos} \end{aligned} \tag{8}$$

where Δw_{ij}^{pre} and Δw_{ij}^{pos} are the values updated in the connection weight between the ith input unit and jth output unit of the pre-, and post-context layers respectively, α is the learning rate which was fixed to 0.0001, t_j is the target value that the jth output unit has to learn, o_j^{pre} and o_j^{pos} are the values of the jth output unit of the pre-/post-context output layers, x_i^{pre} and x_i^{pos} are the input values of the ith input unit of the pre-, and post-context input layers respectively.

After the output activation of the pre-, and post-context (i.e. o_j^{pre} and o_j^{pos}) inputs are calculated from the context input layers individually, the two are multiplied unit-wise to be used as the input for the paradigmatic layer. The multiplication is conducted to capture the activation of words that are associated to both the pre-, and post-context. Therefore, as more contexts are associated (learned) to the target word the multiplication serves as a probabilistic AND function. Words that are highly associated with both pre- and post-contexts have a higher activation, whereas words that are associated only to pre- or post-context do not activate. Also, for computational efficiency, only the top 100 activation units of the multiplied result are retained while other units are set to zero. Then, the activation pattern of the paradigmatic layer was trained to learn the target word at the target output layer using the delta rule. In other words, only the top 100 words that have the strongest association within a particular context were selected to be associated together in a paradigmatic fashion. The weights in the paradigmatic network were updated for every target word in an utterance, and the initial values of all connection weights are set to zero.

Initially, activations in the paradigmatic layer are zero because the weights between the context layers and the paradigmatic layer are zero. Consequently, no learning occurs between the paradigmatic and target output layers. As learning progresses, context to paradigmatic weights are learned, which results in an increased of paradigmatic activations. As a result, learning between the paradigmatic layer and the target output layer gradually accelerates. This mechanism naturally captures the late onset of paradigmatic associations compared to the syntagmatic associations (Brown & Berko, 1960). Furthermore, the same mechanism can be used to explain why adults tend to rely on taxonomic relatedness when extending labels, whereas children do not.

11.1.4. Testing

After training, the model was activated by the word lists used in the behavioral experiments, and fit to the behavioral data. Since we assume that both syntagmatic and paradigmatic associations may contribute to word learning (as long as both types of associations are formed), the words are activated through the input layer in the syntagmatic network, and through the paradigmatic layer in the paradigmatic network. If paradigmatic associates have yet to be formed as we believe is the case with children, there will be no contribution from the paradigmatic network, with the entire contribution coming from the syntagmatic network.

Moreover, to equate the activation level across different experimental conditions, which have different numbers of words, the input activation is normalized:

$$\begin{aligned} o_j^{syn} &= \sum_{i=1}^{n} (w_{ij}^{syn} \cdot x_i^{syn})/k \\ o_j^{par} &= \sum_{i=1}^{n} (w_{ij}^{par} \cdot x_i^{par})/k \end{aligned} \tag{9}$$

where o_j^{syn} and o_j^{par} are the values of the jth output unit in the syntagmatic and paradigmatic network respectively, w_{ij}^{syn} is the connection weight between the ith input unit to the jth output unit in the syntagmatic network, w_{ij}^{par} is the connection weight between ith paradigmatic unit to the jth target output unit in the paradigmatic network, x_i^{syn} is the value of the ith unit in the input layer of the syntagmatic network, x_i^{par} is the value of the ith unit in the paradigmatic layer of the paradigmatic network, n is the total number of input units, and k is the number of words in a given list. From the output of each network the total summed activation for animal units and artifact units are calculated respectively, excluding the activation of the test words. Then, based on the summed activation values, the probability of an animal response for each network (i.e. P_{syn}, and P_{par}) is derived by using a softmax function

$$\begin{aligned} P_{syn} &= \frac{exp\left(\lambda_{syn} \cdot o_{ani}^{syn}\right)}{exp(\lambda_{syn} \cdot o_{ani}^{syn}) + exp(\lambda_{syn} \cdot o_{art}^{syn})} \\ P_{par} &= \frac{exp\left(\lambda_{par} \cdot o_{ani}^{par}\right)}{exp(\lambda_{par} \cdot o_{ani}^{par}) + exp\left(\lambda_{par} \cdot o_{art}^{par}\right)} \end{aligned} \tag{10}$$

where o_{ani}^{syn} and o_{art}^{syn} are the summed activation of the animal and artifact units respectively from the output layer of the syntagmatic network, o_{ani}^{par} and o_{art}^{par} are the summed activation of the animal and artifact units from the target output layer of the paradigmatic network. λ_{syn}, and λ_{par} are free parameters that normalize the relative strength of the animal and artifact activations from each network when transforming the activation strength into decision probability. The two parameters are optimized to fit the behavioral data. Finally, the proportion of animal response (P) is calculated by averaging the proportion of animal activation from the two networks.

$$P = \frac{P_{syn} + P_{par}}{2} \tag{11}$$

11.2. Model simulation

The model was trained on the whole corpus for each epoch. Since each epoch represented a certain amount of learning or experience, the child data should be best captured after a few epochs, while the adult data should be better captured once more epochs have been simulated. To search the number of epochs that would simulate the data best, the child and adult data were each simulated after different number of epochs. A single set of lambda values (i.e., decision parameters λ_{syn}, and λ_{par}) was used for a given child/adult epoch pair. Therefore the number epochs were responsible for the prediction difference between the child and adult performance, and not by the lambda values. For computational efficiency, training was conducted for 50 epochs. The two lambda values (λ_{syn}, and λ_{par}) in equation 5 were optimized for each epoch pair using root-mean-square error (RMSE) as an error term. Fig. 15 depicts the RMSE and correlation coefficient of the simulated values for all epoch pairs. The epoch pair that fits the data best was after the 3rd epoch for the child data and 50th epoch for the adult data (RMSE = 0.102, r = 0.84) using λ_{syn} = 83.02, and λ_{par} = 118,000. These results are of primary importance: they are consistent with the ideas that (a) syntagmatic associations emerge before paradigmatic ones and (b) the former may be necessary for the emergence of the latter ones. In addition, the model captured the results of the 16 experimental conditions

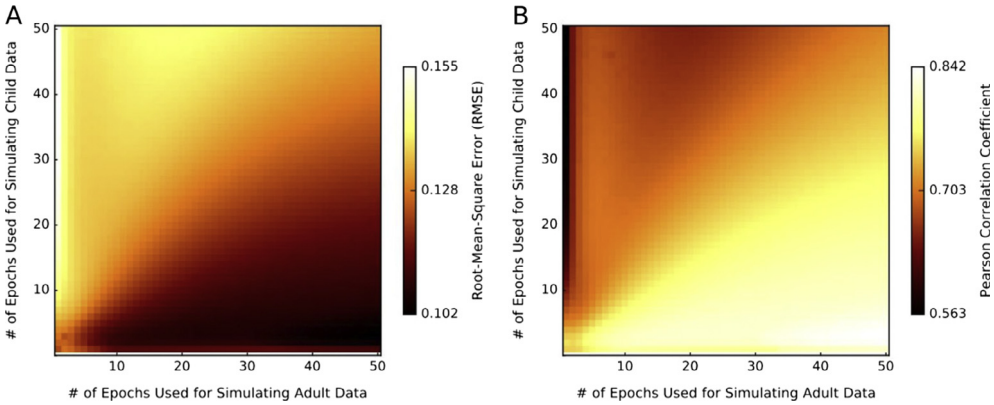

Fig. 15. Root-Mean-Square Error (A) and Pearson correlation coefficient (B) when fitting the data with different epoch pairs.

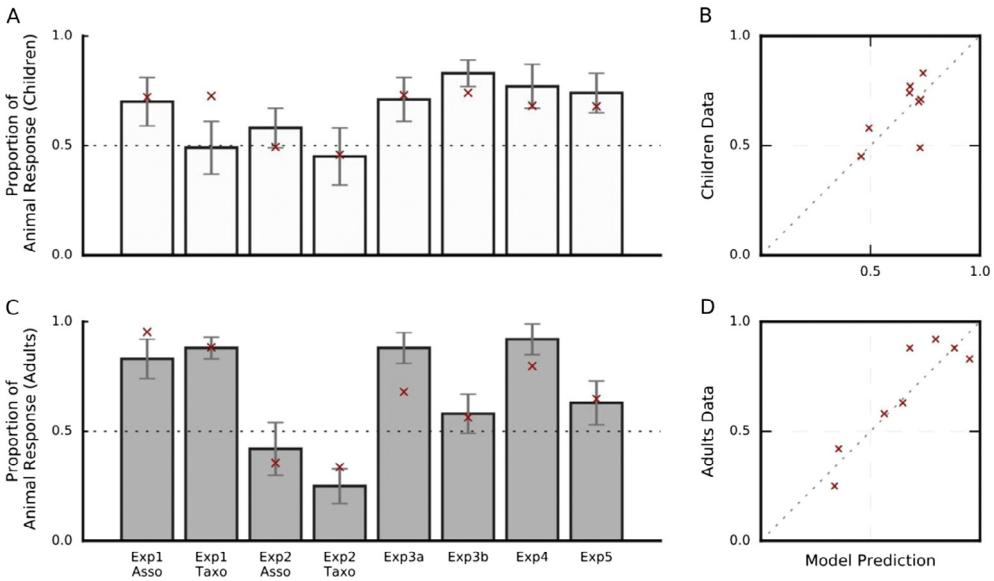

Fig. 16. Simulation results of the best fitting model which used the 3rd epoch for the child data, and the 50th epoch for the adult data. Behavioral data (i.e., proportion of animal response) from the 16 experimental conditions are presented in bars and the values simulated by the model are presented as red crosses. (A) Children's data (presented in bars) and model predictions (in red crosses), (B) Children's data (y-axis) and model prediction (x-axis) presented on a scatter plot, (C) Adults' data (presented in bars) and model predictions (in red crosses), (D) Adults' data (y-axis) and model prediction (x-axis) presented on a scatter plot. Error bars represent ± 1 *SEM*.

remarkably well with only three free parameters, which are the learning rate (identical for the two networks) and two lambda scaling parameters (see Fig. 16).

The proposed model captures the pattern of data and, more importantly, it captures the developmental changes that transpired across the experiments. The results of modeling revealed that a greater number of epochs of training was needed to learn paradigmatic than syntagmatic associations, thus suggesting that more time and experience is needed for learning paradigmatic associations. Therefore, the model captured the fact that syntagmatic associations appear early in development, whereas paradigmatic associations are learned only when a large corpus of syntagmatic associations is available.

In sum, the fact that children exhibit evidence of a syntagmatic to paradigmatic shift in the production of free associations has been known for a long time (e.g., Nelson, 1977, for a review). However the reason for this transition and an underlying mechanism have remained unknown. We propose a solution to this problem that is based on three ideas. First, we propose a mechanistic account of the shift suggesting that syntagmatic associations along with the accumulation of a large corpus of experience with sentences is necessary for the syntagmatic-to-paradigmatic shift to occur. Second, we propose that such shift does not indicate a replacement of syntagmatic associations by paradigmatic ones: whereas young children form mostly syntagmatic associations, older children and adults form both types of association. Finally, we suggest that the proposed mechanistic account of the shift can explain developmental differences in the forward associative strengths (FAS) and subsequently in word learning from context.

12. General discussion

12.1. Summary of findings

A number of important findings stem from the reported experiments. In Experiment 1, children and adults were presented with Associative and Taxonomic animal lists. The summed FAS for children was high for the Associative list and low for the Taxonomic list, whereas for adults it was high for both lists. Children successfully learned the novel animal word in the Associative condition, whereas there was no learning in the Taxonomic condition. In contrast, adults learned successfully in both the Associative and the Taxonomic conditions.

In Experiment 2, children and adults were presented with Associative and Taxonomic artifact lists. For children, the summed FAS of the Associative list was substantially lower than that in Experiment 1, whereas the summed FAS of the Taxonomic list was as low as in Experiment 1. For adults, the summed FAS of either list was lower than that in Experiment 1. Children failed to learn the novel word in either condition, whereas adults exhibited evidence of learning only in the Taxonomic condition.

In Experiment 3, children and adults were presented with substantially shortened Associative animal lists. The associative word list was reduced to 4 words (Experiment 3A) or to 1 word (Experiment 3B), while the summed FAS to the word *animal* in either list remained high. Participants learned successfully in both experiments, with adults exhibiting somewhat weaker learning in Experiment 3B.

And finally, in Experiments 4–5, children and adults were presented with mixed lists. In Experiment 4, the animal associative list was mixed with color words, whereas in Experiment 5 this list was mixed with the artifact associative list. Critically, in both lists, the summed FAS with the word *animal* remained high. In both experiments, children and adults successfully learned the words.

Taken together, these results suggest that words could be learned from the context, and that the summed FAS of the list predicts learning. These results have important theoretical implications, and we discuss these implications in the next section.

12.2. Mechanisms of word learning

The fact that across the reported experiments summed FAS predicted learning (and it did it better than the competitors) supports the importance of associative mechanisms in word learning. At the same time, these findings are difficult to reconcile with either variant of the inference approach to word learning that we considered. Recall that the XTB has difficulty accounting for results of children's word learning, faring somewhat better with adults (see Fig. 10). TOPICS did better than XTB, but it exhibited a strong animal bias, thus failing to account for adult learning in artifact conditions or children's failure to learn in the Taxonomic animal condition. Finally, FAS-based predictions were accurate for most of experiments with children, and these were comparable with the other two models for adults (see Fig. 12).

It is also not clear how these findings could be explained by the social reasoning approach. Recall that according to this view, understanding the intent of the speaker is a critical component of word learning. Therefore, word learning should be the best in those conditions where the speaker's intent is the clearest (i.e., in the Taxonomic condition of Experiment 1) and it should be the worst where the speaker's intent is the least clear (i.e., when learner is presented with a mixed list, such as in Experiments 4 and 5). However, the opposite was observed – children were more likely to learn a new word in the latter conditions than in the former condition.

Finally, it is not clear how the idea that children hold the taxonomic assumption could account for these findings. If word learning is driven by the assumption that words refer to taxonomic kinds, then the condition where every word refers to the same taxonomic kind (i.e., the Taxonomic condition of Experiment 1) should provide better support for word learning than the condition where none of the words refer to the taxonomic kind (i.e., Experiments 3A and 3B).

At the very minimum, the results of the reported experiments indicate that rational inference may not be the right framework for learning words from a context and that appropriate meaning could be automatically activated by the context. Under this account, word meaning could be initially gleaned from the linguistic context, in which these words occur. This meaning is however rough and imprecise – its generalization is driven by syntagmatic associations – by the words that appear in a close temporal proximity in a sentence. Two major changes occur in the course of development. First, the words that co-occur (and thus are likely to be related thematically) become associated. And second, as children get exposed to the same word in a variety of contexts, words that occur in similar contexts (and thus are likely to be related taxonomically) become associated. These different (i.e., syntagmatic and paradigmatic) associations impose different (i.e., thematic and taxonomic) constraints on word learning.

The fact that paradigmatic associations emerge later in development than syntagmatic associations has been known for some time (Brown & Berko, 1960; Ervin-Tripp, 1970; see also Nelson, 1977, for a review). In particular, a number of researchers have observed that when presented with a free association task, adults tend to respond with a word that is taxonomically related to the target word (e.g., *night* --> *day*), whereas young children tend to respond with a thematically related word, the one that frequently follows or precedes the target word (e.g., *night* --> *dark*). Although this developmental asynchrony (often referred to as "the syntagmatic to paradigmatic shift") has been extensively discussed in previous work, there has been no

explicit developmental account for this shift. We believe that our model offers such an account and we discuss it in the next section.

12.3. What develops? Evidence from computational modeling

Why is there a syntagmatic to paradigmatic "shift" -- a transition from only syntagmatic associations in children to a combination of syntagmatic and paradigmatic associations in adults? To answer this question, we developed and tested a new model providing a mechanistic account of syntagmatic associations, of paradigmatic associations, and of the developmental transition from word learning in children to word learning in adults. The fundamental idea is that the transition from syntagmatic to paradigmatic associations is experience-based: paradigmatic associations emerge as a result of experience with order-dependent syntagmatic associations. Children first learn the statistics of co-occurrences -- syntagmatic associations among words that appear in the same order dependent contexts, such as *lion* and *zoo*, *tiger* and *zoo*, and *animal* and *zoo*. On the basis of multiple syntagmatic associations, children start learning paradigmatic associations -- associations among words that share the same syntagmatic associations. Paradigmatic associations are second-order and they link words that play a similar role in the sentence, such as *lion*, *tiger*, and *animal*. In other words, substantial experience with language is necessary to start to form paradigmatic associations, the experience that adults have and young children miss.

The model provides evidence for this experience-based developmental change by demonstrating that by changing the number of learning epochs during training, it can account for differences in word learning between children and adults. This was possible since the paradigmatic network required additional experience to learn compared to the syntagmatic network. Therefore, initially only the syntagmatic network affects the model's performance, which was sufficient to account for the child data. In contrast, to capture the adult data, the additional learning in the paradigmatic network was required, which developed later in training. This simple idea accounts for results of the 16 reported conditions, without assuming any inference or hypothesis testing mechanisms (see Fig. 16). This idea also offers a mechanistic account of syntagmatic to paradigmatic shift.

This developmental idea is not unreasonable: There is empirical evidence supporting this possibility (McNeill, 1963). In the McNeill's (1963) experiment, adult participants were presented with sentences that included a nonsense adjective (e.g., *Koj*) and a nonsense count noun (e.g. *Maf*). In addition, sentences that matched exactly, but with the second nonsense noun replaced by a difference nonsense noun (e.g. *Zon*) were included. Consequently, *Koj* appeared directly with both *Maf* and *Zon*, but *Maf* and *Zon* never appeared together. Participants were then presented with a free association task after 20, 40, or 60 trials of training. Initially, after 20 trials of training, participants produced predominantly syntagmatic associations (i.e., when presented with *Maf*, they responding *Koj*). However, after 60 trials of training they started responded both syntagmatically (as above) as well as paradigmatically (i.e., when presented with *Maf*, they responded *Zon*). These findings support the idea that paradigmatic associations take longer to emerge (and perhaps they require some experience with syntagmatic associations).

The proposed model is capable of offering a mechanistic account of learning words from context and of syntagmatic to paradigmatic shift in development and, as we discuss below it offers solutions to the mapping and generalization problems. At the same time, additional research is needed to examine the ability of the model to account for the multiple "benchmark" phenomena in word learning.

12.4. Consequences for the mapping and generalization problem

A critical advantage of the syntagmatic/paradigmatic approach is that it provides possible solutions to both the mapping and generalization problems:

12.4.1. Mapping problem

Recall Quine's (1960) observation that the word *gavagai* uttered while pointing to a rabbit has an infinitely large number of potential mappings, including the rabbit, its parts, its texture, its method of locomotion, or the speaker's attitude towards the rabbit. In this ostensive context and in the absence of any previous linguistic experience, we would argue that the meaning is indeterminate. However, the linguistic contexts in which one hears the word *gavagai* may help quickly isolate the relevant meaning. If the learner hears the utterance "Mom fed the gavagai" then order-dependent syntagmatic associations formed as a consequence of exposure to previous similar utterances such as "Dad fed the cats" or "Sofie fed the hamster" will generate the prediction that {cats, hamster} will appear. When *gavagai* ultimately does appear it will become paradigmatically associated with {cats, hamster} and therefore licensed to be used in the ways that these words are. One does not typically hear "Mom fed the foot" or "Mom fed the soft" or "Mom fed the hopping". Note, there is no circularity as it is not necessary that the meaning of "fed" or "the" exist. It is enough that the syntagmatic associations have been formed between whole word perceptual representations. The meaning of the word then emerges as a consequence of the interaction of the syntagmatic and paradigmatic associations between words and is therefore inherently contextual.

Note also that the notion of paradigmatic association does not need to be restricted to linguistic tokens. The essence is that an association is formed between a predicted perceptual representation and the perceptual representation that actually occurs at that time. The pointing operation in ostensive settings is a form of visual expression that is interwoven with the linguistic stream. If one is naming an object then one points and says the name. So, one might experience "<point> cats", and

"<point> hamster" and therefore predict {cats, hamster} when the pointing operation occurs. By way of contrast, the fondling gesture might predict a different set of words. Having experienced "<fondle> rough" and "<fondle> soft", when experiencing "<fondle> gavagai" the learner may not infer that gavagai means rabbit, but rather that gavagai means fluffy. Holding or riding might predict a different set of words and so on.

This approach to the mapping problem is similar to the idea proposed by Goldstone and Rogosky (2002). According to these researchers, there are two sources of constraints on word meaning, one stemming from the semantic domain (i.e., interrelatedness of words) and another from the external world, including the visual world. Therefore, in the absence of semantic constraints (e.g., when the word *Gavagai* is heard for the first time), it is likely that only the constraints of the external world would be applied. If indeed objects (rather than colors or textures) are units of attention (e.g., Kanwisher & Driver, 1992), then it would be natural to map *Gavagai* to the whole object. Under this interpretation, the whole object constraint may come from experience with allocating visual attention. Although this idea seems similar to the whole object assumption (Markman, 1990), it differs in an important way. The whole object assumption is construed as an *a priori* and language-specific constraint, whereas the described above constraint is domain-general and potentially experience-based. Therefore, in contrast to the language-specific whole-object assumption, the latter constraint, because of its domain-general nature, may affect other aspects of learning. This is a preliminary idea and further research is needed for testing it and better understanding how and under what conditions this constraint could be relaxed.

12.4.2. Generalization problem

There is also substantial constraint imposed by the linguistic context that can be used to determine the scope with which a word should be extended. In the ostensive context, the word *gavagai* has an infinitely large number of potential extensions: this particular rabbit, some rabbits, all rabbits, rabbits and dogs, all animals, all solid objects, all things, etc. However, the linguistic contexts in which it appears constrain the appropriate scope rapidly. If the learner hears "Mom fed the gavagai" then paradigmatic associations will be formed to a broader class of animals{cats, hamster, horses}, as above. In contrast, if one hears "Sofie cuddled our gavagai", having heard only "Sofie cuddled our hamster" the set narrows to {hamster} which in the context of this family licences usages the refer to a particular rabbit as opposed to all rabbits or all animals.

For expository purposes, we have represented paradigmatic associations as simple sets. We assume, however, that the membership in the set of paradigmatic associates would be graded as a consequence of the strength of prediction afforded by the linguistic context. The set of associates generated by "Mom fed the gavagai" would be primarily determined by the order-dependent syntagmatic associates of "fed" and "the". The predictions that can be made based on the presence of "the" will be many but weak. The predictions afforded by the presence of "fed" will be fewer and stronger. Embedded within the syntagmatic/paradigmatic approach is a notion that is different from, but akin to the size principle proposed by XTB. From the single exposure to "Mom fed the gavaigai", it is possible to learn that a gavagai is both a thing and an animal, suitable for use in contexts such as "The gavagai was destroyed" and "The gavagai was killed" (the latter being preferable).

13. Conclusion

Word learning is a notoriously difficult problem, and it has been argued that associative accounts are unable to explain the phenomenon. Moreover, whereas most words in the lexicon are learned from context, much of research on word learning has been focused on learning by ostension. In this research, we (a) examined how children learn words from context and (b) demonstrated the role of associations in this process. We suggest that learning words from context is initially driven by syntagmatic associations, whereas later in development both syntagmatic and paradigmatic associations contribute to word learning. Therefore, the syntagmatic to paradigmatic shift can explain differences in word learning between children and adults. Finally, our computational model offers a mechanistic account of the syntagmatic to paradigmatic shift, while providing important insights about learning words from context and developmental changes in this process. In sum, the reported results point to a powerful associative mechanism of word learning. At the same time, the reported findings present interesting challenges for the inference accounts of word learning.

Author note

This research is supported by the National Institutes of Health Grants R01HD078545 and P01HD080679 to Vladimir M. Sloutsky. We thank Chris Robinson, Anna Fisher, and members of the Cognitive Development Lab for helpful comments. Address correspondence to Vladimir M. Sloutsky (sloutsky.1@osu.edu), Department of Psychology, The Ohio State University, Columbus, OH 43210.

Appendix A

As the word set that we employed was not identical to that employed by Xu and Tenenbaum (2007), it was necessary to create a new hierarchical clustering using our terms in order to apply their model. Response probability vectors were calculated for each of the nouns based on the De Deyne free association norms (De Deyne, Navarro, & Storms, 2012). When the De Deyne norms were collected participants were required to produce three responses. The probability vectors we calculated

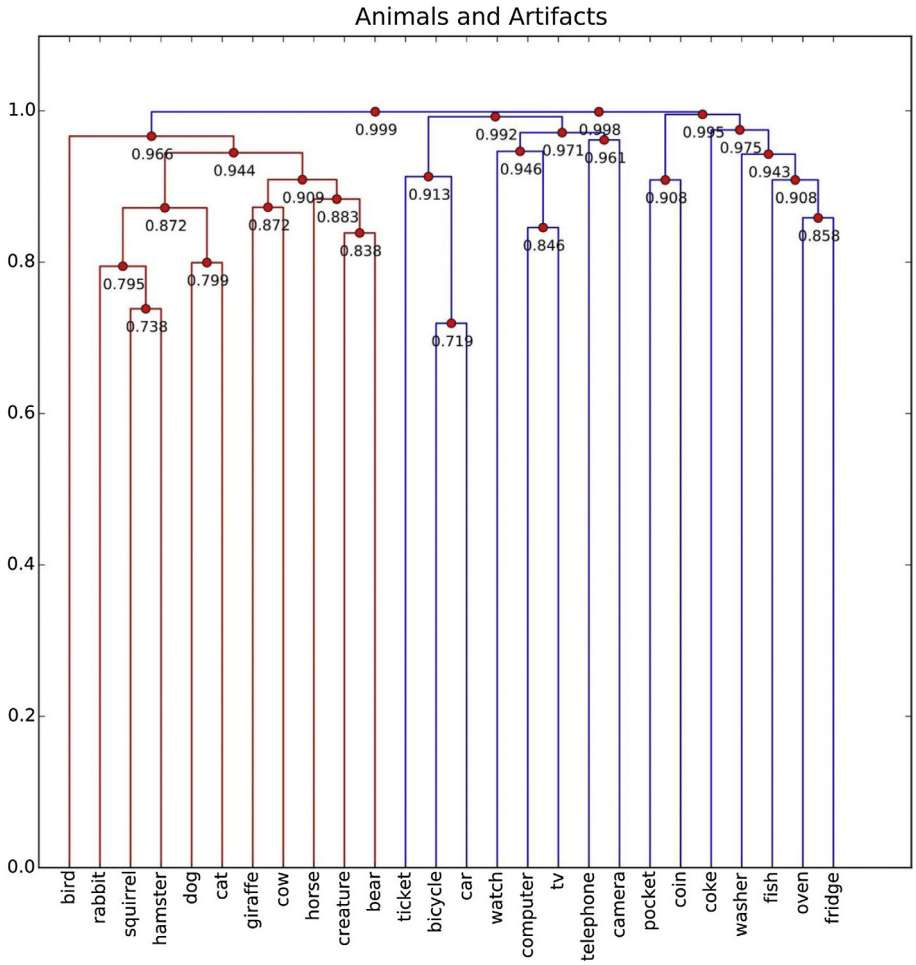

Fig. A1. The hierarchical cluster analysis of the nouns in our experiment.

were based on all three responses. Fig. A1 shows the single linkage hierarchical clustering solution based on the square root of the Jenson Shannon Divergence (JSD) between these vectors:

$$JSD(P\|Q) = \frac{D(P\|M)}{2} + \frac{D(Q\|M)}{2} \tag{A1}$$

where

$$M = \frac{P+Q}{2} \tag{A2}$$

and D is the Kulback Leibler divergence defined as:

$$D(P\|Q) = \sum_i P_i \ln(P_i/Q_i) \tag{A3}$$

Note the square root of the Jenson Shannon divergence is employed because it is a metric.

The major division is between animals and artifacts, while many of the subclusters also make intuitive sense (rabbits, squirrels and hamsters are clustered together; as are cats and dogs; bicycles and cars and TVs and computers). The one exception to the taxonomic character of the result is the word fish which appears near oven and fridge, presumable because it is a food. In our simulations we assumed that fish appeared on the animal cluster.

Appendix B

See Fig. B1.

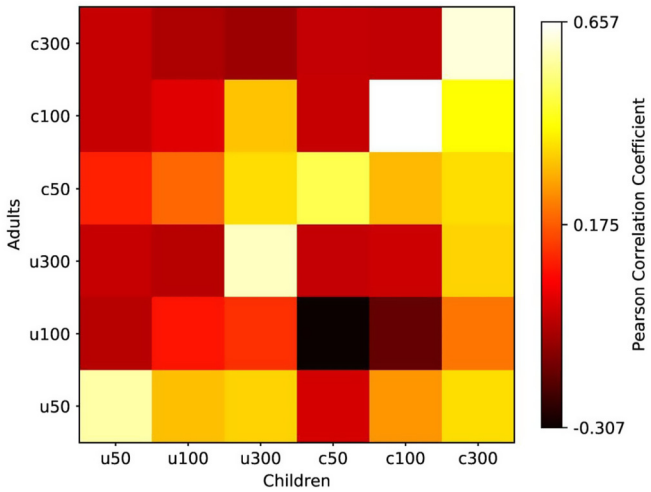

Fig. B1. Model fit of the TOPICS model with different document size and number of topics. The label on the axis represents document size (i.e., 'u' for utterance, and 'c' for conversation), and number of topics (i.e., 50, 100, and 300). As described in the main text, each model with a different document size and number of topics was evaluated on how well it fit the children and adult data separately. The best fitting models are, in terms of correlation value, conversation with 100 topics (i.e., c100) for both adults and children ($r = .657$).

Appendix C

See Table C.

Table C
Suffix used for filtering the corpus.

-'s	-ing	-iest	-icide	-less	-ic
-s'	-ier	-est	-ience	-iful	-al
-iness	-er	-ist	-ence	-ful	-ily
-ness	-or	-ism	-iance	-iant	-ably
-ution	-ate	-ish	-ance	-ant	-ibly
-sion	-acy	-ive	-able	-rrent	-ly
-ion	-tial	-ility	-ible	-iment	-ies
-ied	-ian	-ity	-ious	-ment	-es
-ed	-hood	-ize	-ous	-en	-s

References

Akhtar, N. (2002). Relevance and early word learning. *Journal of Child Language, 29*, 677–686.
Akhtar, N., & Tomasello, M. (2000). The social nature of words and word learning. In R. M. Golinkoff & K. Hirsh-Pasek (Eds.), *Becoming a word learner: A debate on lexical acquisition* (pp. 115–135). Oxford: Oxford University Press.
Baldwin, D. A. (1991). Infants' contribution to the achievement of joint reference. *Child Development, 62*, 875–890.
Baldwin, D. A. (1993). Early referential understanding: Infants' ability to recognize referential acts for what they are. *Developmental Psychology, 29*, 832.
Bloom, L. (1973). *One word at a time: The use of single word utterances before syntax*. The Hague: Mouton.
Bloom, P. (2000). *How children learn the meanings of words*. Cambridge, MA: MIT Press.
Brown, R. (1957). Linguistic determinism and the part of speech. *Journal of Abnormal and Social Psychology, 55*, 1–5.
Brown, R., & Berko, J. (1960). Word association and the acquisition of grammar. *Child Development, 31*, 1–14.
Colunga, E., & Smith, L. B. (2005). From the lexicon to expectations about kinds: A role for associative learning. *Psychological Review, 112*, 347–382.
De Deyne, S., Navarro, D., & Storms, G. (2012). Better explanations of lexical and semantic cognition using networks derived from continued rather than single word associations. *Behavior Research Methods, 45*, 480–498.
Dennis, S. (2005). A memory-based theory of verbal cognition. *Cognitive Science, 29*, 145–193.
Elman, J. L. (1990). Finding structure in time. *Cognitive Science, 14*, 179–211.
Elman, J. L. (2009). On the meaning of words and dinosaur bones: Lexical knowledge without a lexicon. *Cognitive Science, 33*, 547–582.
Ervin-Tripp, S. M. (1970). Substitution, context and association. In L. Postman & G. Keppel (Eds.), *Norms of word association* (pp. 383–467). New York: Academic Press.
Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., & Pethick, S. J. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development, 59*, 1–173.
Fisher, C. (1996). Structural limits on verb mapping: The role of analogy in children's interpretations of sentences. *Cognitive Psychology, 31*, 41–81.
Gelman, S. A., & Davidson, N. S. (2013). Conceptual influences on category-based induction. *Cognitive Psychology, 66*, 327–353. http://dx.doi.org/10.1016/j.cogpsych.2013.02.001.

Gleitman, L. R. (1990). The structural sources of verb meanings. *Language Acquisition, 1*, 3–55.
Goldstone, R. L., & Rogosky, B. J. (2002). Using relations within conceptual systems to translate across conceptual systems. *Cognition, 84*, 295–320.
Goodman, J. C., McDonough, L., & Brown, N. B. (1998). The role of semantic context and memory in the acquisition of novel nouns. *Child Development, 69*, 1330–1344.
Griffiths, T. L., Steyvers, M., & Tenenbaum, J. B. (2007). Topics in semantic representation. *Psychological Review, 114*, 211–244. http://dx.doi.org/10.1037/0033-295X.114.2.211.
Hockett (1960). The Origin of Speech. *Scientific American, 203*, 88–111.
Jenkins, G., Samuelson, L. K., Smith, J. R., & Spencer, J. P. (2015). Non-Bayesian noun generalization in 3-5-year-old children: Probing the role of prior knowledge in the suspicious coincidence effect. *Cognitive Science, 39*, 268–306.
Jones, M. N., & Mewhort, D. J. K. (2007). Representing word meaning and order information in a composite holographic lexicon. *Psychological Review, 114*, 1–37.
Kanwisher, N., & Driver, J. (1992). Objects, attributes, and visual attention: Which what, and where. *Current Directions in Psychological Science, 1*, 26–31.
Landau, B., & Gleitman, L. R. (1985). *Language and experience: Evidence from the blind child*. Cambridge, MA: Harvard University Press.
Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review, 104*, 211–240.
Luche, C. D., Durrant, S., Floccia, C., & Plunkett, K. (2014). Implicit meaning in 18-month-old toddlers. *Developmental Science, 17*, 948–955.
MacWhinney, B. (2000). *The CHILDES Project: Tools for analyzing talk*. 3rd Edition. Vol. 2: The Database. Mahwah, NJ: Lawrence Erlbaum Associates.
Markman, E. M. (1989). *Categorization and naming in children*. Cambridge, MA: MIT Press.
Markman, E. M. (1990). Constraints children place on word meanings. *Cognitive Science, 14*. 57–7.
Markman, E. M., & Hutchinson, J. E. (1984). Children's sensitivity to constraints on word meaning: Taxonomic versus thematic relations. *Cognitive Psychology, 8*, 561–577.
Mayor, J., & Plunkett, K. (2010). A neurocomputational account of taxonomic responding and fast mapping in early word learning. *Psychological Review, 117*, 1–31. http://dx.doi.org/10.1037/a0018130.
McMurray, B., Horst, J. S., & Samuelson, L. K. (2012). Word learning emerges from the interaction of online referent selection and slow associative learning. *Psychological Review, 119*, 831–877. http://dx.doi.org/10.1037/a0029872.
McNeill, D. (1963). The origin of associations within the same grammatical class. *Journal of Verbal Learning and Verbal Behavior, 2*, 250–262. http://dx.doi.org/10.1016/S0022-5371(63)80091-2.
Nagy, W. E., Herman, P. A., & Anderson, R. C. (1985). Learning words from context. *Reading Research Quarterly, 20*, 233–253.
Naigles, L. (1990). Children use syntax to learn verb meanings. *Journal of Child Language, 17*, 357–374.
Nelson, K. (1977). The syntagmatic-paradigmatic shift revisited: A review of research and theory. *Psychological Bulletin, 84*, 93–116.
Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (2004). The University of South Florida free association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments, & Computers, 36*, 402–407.
Preston, A. R., Shrager, Y., Dudukovic, N. M., & Gabrieli, J. D. E. (2004). Hippocampal contribution to the novel use of relational information in declarative memory. *Hippocampus, 14*, 148–152.
Quine, W. V. O. (1960). *Word and object*. Cambridge, MA: MIT Press.
Redington, M. (1998). Distributional information: A powerful cue for acquiring syntactic categories. *Cognitive Science, 22*(4), 425–469.
Regier, T. (2005). The emergence of words: Attentional learning in form and meaning. *Cognitive Science, 29*, 819–865.
Robinson, C. W., & Sloutsky, V. M. (2007). Visual processing speed: Effects of auditory input on visual processing. *Developmental Science, 10*, 734–740. http://dx.doi.org/10.1111/j.1467-7687.2007.00627.x.
Robinson, C. W., & Sloutsky, V. M. (2010). Development of cross-modal processing. *Wiley Interdisciplinary Reviews: Cognitive Science, 1*, 135–141. http://dx.doi.org/10.1002/wcs.12.
Samuelson, L. K., & Smith, L. B. (1999). Early noun vocabularies: Do ontology, category structure and syntax correspond? *Cognition, 73*, 1–33.
Siegler, R. S., & Svetina, M. (2006). What leads children to adopt new strategies? A Microgenetic/cross-sectional study of class inclusion. *Child Development, 77*, 997–1015.
Siskind, J. M. (1996). A computational study of cross-situational techniques for learning word-to-meaning mappings. *Cognition, 61*, 39–91.
Soja, N. N. (1992). Inferences about the meanings of nouns: The relationship between perception and syntax. *Cognitive Development, 7*, 29–45.
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.
Stevers, M., & Griffiths, T. (2007). Probabilistic topic models. In T. Landauer, D. McNamara, S. Dennis, & W. Kintsch (Eds.), *Handbook of latent semantic analysis* (pp. 424–440). Mahwah, NJ: Laurence Erlbaum.
Wilson, M. (1988). MRC Psycholinguistic Database: Machine-usable dictionary, version 2.00. *Behavior Research Methods, Instruments, & Computers, 20*, 6–10.
Winer, G. A. (1980). Class-inclusion reasoning in children: A Review of the empirical literature. *Child Development, 51*(2), 309–328.
Xu, F., & Tenenbaum, J. B. (2007). Word learning as Bayesian inference. *Psychological Review, 114*, 245–272.
Yoshida, H., & Smith, L. B. (2003). Shifting ontological boundaries: How Japanese-and English-speaking children generalize names for animals and artifacts. *Developmental Science, 6*, 1–17.
Yu, C., & Smith, L. B. (2007). Rapid word learning under uncertainty via cross-situational statistics. *Psychological Science, 18*, 414–420.
Yu, C., & Smith, L. B. (2012). Modeling cross-situational word-referent learning: Prior questions. *Psychological Review, 119*, 21–39. http://dx.doi.org/10.1037/a0026182.
Zeithamova, D., & Preston, A. R. (2010). Flexible memories: Differential roles for medial temporal lobe and prefrontal cortex in cross-episode binding. *Journal of Neuroscience, 30*, 14676–14684. http://dx.doi.org/10.1523/JNEUROSCI.3250-10.2010.