# Early-Talker and Late-Talker Toddlers and Networks Show Different Word Learning Biases

**Eliana Colunga (eliana.colunga@colorado.edu)**
Department of Psychology and Neuroscience, 345 UCB
Boulder, CO 80309-0345 USA

**Clare E. Sims (clare.holtpatrick@colorado.edu)**
Department of Psychology and Neuroscience, 345 UCB
Boulder, CO 80309-0345 USA

**Abstract**

In typical development, word learning goes from slow and laborious to fast and seemingly effortless. Typically developing 2-year-olds are so skilled at learning noun categories that they seem to intuit the whole range of things in the category from hearing a single instance named – they are biased learners. This is not the case for children below the 20th percentile on productive vocabulary (*late talkers*). This paper looks at the individual vocabularies and word-learning biases of late- and early-talking toddlers. Experiment 1 shows that neural networks trained on the vocabularies of individual late talkers learn qualitatively different biases than those trained on early talker vocabularies. Experiment 2 confirms the novel predictions made by the simulations about word learning biases in late- vs. early-talking children. The implications for diagnosis and intervention are discussed.

**Keywords:** Late talkers; early talkers; computational models; neural networks, word learning.

## Introduction

There is extraordinary variability in the vocabularies of very young children. A two-year-old in the lower 10th percentile may produce around 10 words whereas a two-year-old in the top 10th percentile will produce well over 300 (Fenson, Dale, Reznick, Thal, Bates, Hartung, Pethick, & Reilly, 1993). In general, the course of word learning proceeds from slow, effortful learning of nouns and of the range of things that belong in a category, to very rapid learning of object names. Indeed, typically developing 2-year-olds are so skilled at learning new nouns that they seem to intuit the whole range of things in a named category from a single naming experience. This is not necessarily the case for children below the 15th-20th percentile on productive vocabulary, or *late talkers*. Why do some children learn words quickly and early and others learn words slowly, maybe even showing effects that persist into adolescence? This paper looks at two possible contributing, and interrelated, factors: noun vocabulary composition and word learning biases

The evidence suggests that children are skilled noun learners because they know about the different kinds of properties that are relevant for categorizing different kinds of things. Typically-developing children show word learning biases that are specific to different kinds of things: they generalize names for solid objects by shape and names for non-solid substances by material (e.g., Soja, Carey, & Spelke, 1991).

The evidence also suggests that children learn how to learn nouns – and learn how different kinds of properties are relevant for different kinds of things – as a consequence of learning names for things. Each noun the child learns appears to teach the child something general about how to learn new nouns that name things of that same kind, and critically, at the same time, this learned general knowledge constrains and facilitates the types of nouns the child will learn next. Through computational models and a study with toddlers, we show that even before they turn 2, late- and early-talker toddlers show different word learning biases.

### Vocabulary composition and word learning biases

The relationship between vocabulary composition and word learning biases has been typically characterized in one of two ways: abstract knowledge guides, facilitates and indeed allows word learning (e.g., Soja et al, 1991; Gelman & Bloom, 2000) or the words that have been learned give rise to, create, and in fact constitute generalized knowledge about word learning (e.g., Colunga & Smith, 2005, Samuelson, 2008). We would like to bypass the debate on whether word-learning biases are the egg to the vocabulary chicken or the other way around and focus instead on the interrelationship between these two factors.

In the domain of names for objects and substances, and in typical development, vocabulary structure and abstract knowledge in the form of kind-specific generalizations appear to be tightly coupled. First, the tendency to attend to shape in the specific context of naming artifacts emerges as children learn nouns, becoming particularly robust around the time children have between 50 to 150 nouns in their productive vocabulary (Gershkoff-Stowe & Smith, 2004). Second, the order of development of these word learning biases reflects the statistical structure of early noun vocabularies, (Samuelson & Smith, 1999; Colunga & Smith, 2005). Third, changing 17-month-olds' vocabulary composition by intensively teaching them names for artifacts yields an early bias to generalize names for artifacts by shape *and accelerates learning of object names outside of the lab*, causing a dramatic increase in vocabulary size for children in the experimental training group but not for those in the control groups (Smith, Jones, Landau, Gershkoff-Stowe & Samuelson, 2002). Fourth, computational models

trained on the structure of the average 30-month-old vocabulary, show word learning biases like those of young children when processing new objects (Colunga & Smith, 2005), and further the structure of the training set affects subsequent training, facilitating the learning of some sorts of categories but hindering others (Colunga, in prep). Altogether, these results suggest a developmental feedback loop between learning object names, developing biases to attend to the relevant properties for artifacts, and the learning of more object names.

## Late Talkers

Children below the 15th-20th percentile on normative measures of productive vocabulary size, so-called late talkers, are not a homogenous group in terms of their developmental outcomes: some catch up, and some show persistent delays (Rescorla, 2002, Rescorla, Roberts, & Dahlsgaard, 1997). However, Rescorla and colleagues argue against considering late talkers and typically developing children as distinct groups, and argue instead for conceptualizing them in terms of a "language endowment spectrum." Importantly, although there is continuity in vocabulary measures at the group level, the outcome for individual children cannot be accurately predicted on the basis of vocabulary production or comprehension (e.g., Desmarais, Meyer, Bairati & Rouleau, 2008).

The literature briefly reviewed above suggests that, in typical development, the words a child knows and what the child knows about learning words in general go hand in hand, and that learning names for categories of things organized by shape speeds up learning nouns. However, this may not be the case for all children. Unlike typically developing children, late talkers do not systematically extend the name of a novel solid object to other objects that match it in shape, and in fact, in one study, almost half of the late talkers systematically extended the novel name of a solid object to others matching in texture rather than shape (Jones, 2003). Recent evidence suggests that the vocabularies of children of different language abilities may be structured differently (Colunga & Sims, 2011; Beckage, Smith & Hills, 2011). These findings suggest that late talkers may not just limited in their production of object names (the measure that defines them as late talkers) but also deficient in the processes that subserve the acquisition of new words and in their knowledge about those categories. The crucial question, then, is whether these differences in vocabulary composition are differences that matter. Do the different nouns late- and early-talkers know yield different word learning biases? In two experiments we test the relationship between vocabulary composition and word learning biases, first in neural networks (Experiment 1) and then with 1-year-old toddlers in the lab (Experiment 2). For the purposes of this paper we will focus on contrasting the vocabularies of children on the two opposite ends of the spectrum, late talkers and early talkers.

If the differences in vocabulary structure can, to some extent, explain the differences in language ability, we would expect late talker vocabularies to yield different word learning biases than early talker vocabularies. More specifically, we would expect early talker vocabularies to yield word learning biases that would facilitate the learning of a vocabulary structured like the MCDI – highlighting shape similarities for solids and material similarities for non-solids. In contrast, we would expect networks trained on late talkers' vocabularies to generalize more variable word learning biases, and perhaps even biases that would be unhelpful in learning early vocabularies.

# Experiment 1

## Method

**Materials.** The vocabulary measure used was the Bates-MacArthur Communicative Developmental Inventory toddler version (MCDI) both to select children and to measure vocabulary composition. This is a parent checklist that asks parents to indicate the words that their child *produces* and although it is imperfect as a measurement instrument (Fenson, et al, 1994) it appears to be reliable and to be systematically related to children's performances in a variety laboratory measures of word learning, including especially their word-learning biases in the Novel Noun Generalization (NNG) task (e.g., Landau, et al, 1988).

**Participants.** The vocabularies of 15 late talkers and of 15 early talkers were selected out of a pool of 148 parent-filled MCDI forms for children between 18-30 months of age. The criterion for inclusion was that there existed a vocabulary form from a child matching in age to within 5 days in both the late talker and the early talker groups. Late talkers fell under the 25th percentile; early talkers were above the 75th percentile according to the MDCI norms.

The ages for the two language groups ranged from 18.49 months to 28.26 months (M=23.14 and 23.15 for late and early talkers respectively. Vocabulary sizes for the late talker group ranged between 15 and 425 words (M=132.53); for the early talker group vocabulary size was between 158 and 664 words (M=457).

The noun vocabularies for each individual child were characterized by looking at the proportion of nouns they knew for each of the following categories: 1) solid things alike in shape (e.g., spoon), 2) solid things alike in material (e.g., chalk), 3) solid things alike in both shape and material (e.g., penny), 4) non-solid things alike in shape (e.g., bubble), 5) non-solid things alike in material (e.g., milk), 6) non-solid things alike in both (e.g., jeans). Nouns in children's vocabularies were classified as falling in each of these categories according to adult judgments made for each of the nouns in the MCDI reported in Samuelson & Smith, 1999. The training sets were then constructed to mimic the vocabulary composition of each child (see below).

**Architecture.** The computational models are a modified version of the ones Colunga & Smith, 2005. The main difference is that these networks were trained using the Leabra algorithm, an algorithm that combines Hebbian and error driven learning (O'Reilly, 1996), instead of Contrastive Hebbian Learning as in the original simulations.

The networks are organized as follows: Words are represented discretely (as single units) and are input on the Word Layer (Figure 1). Referents are represented as distributed patterns over several dimensions on the Perception Layer. For example, the shape and material of an object (say the roundness of a particular ball and its yellow rubbery material) are represented by an activation pattern along the Perception layer. Solidity and Non-solidity are represented discretely; one unit stands for Solid and another for Non-Solid. Finally, there is a hidden layer that is connected to all the other layers and to itself. These networks have been shown to model performance in an analog of the NNG Task when trained on vocabularies structured as those of the average 30-month-old.

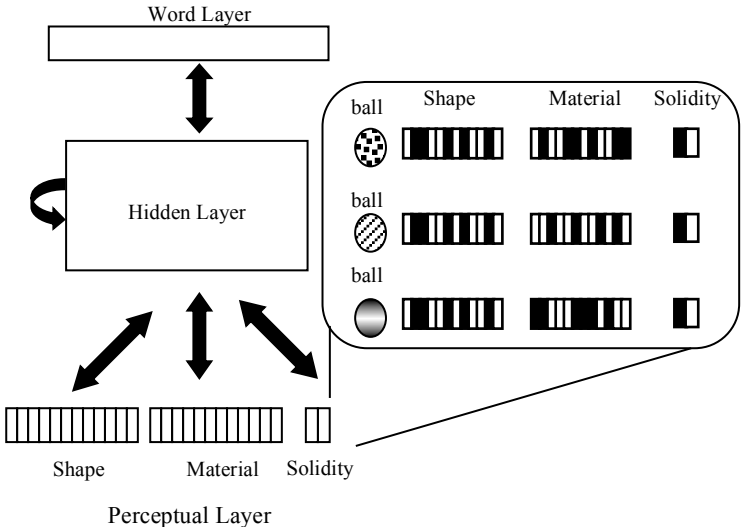


**Figure 1: Architecture of the network and example input patterns.**

**Training.** The networks are trained with categories presenting the same correlational structure as each individual child's noun vocabulary. On each training trial, a word is paired with a referent. The patterns associated with each word are determined by adult judgments of the early noun corpus. For example, adults judged balls to be similar in shape but different in material. To simulate this, we randomly selected an input vector to represent ball shape. Then on individual training trials, we paired that pattern with the label ball and a randomly selected material pattern (Figure 1). We do this for each noun in the training set. Each network was trained in this way for its simulated vocabulary until they reached asymptotic (and near perfect) performance. This part of the simulation is intended to put into the networks the lexical knowledge that the individual child would bring to the laboratory NNG task.

Because we are interested in the consequences of different vocabulary structures regardless of their size, all networks were trained to learn 24 nouns, proportionally structured like their corresponding child's vocabulary. Thus, the only difference between networks were the differences in vocabulary composition for each individual child.

**Testing.** The question is what sort of word learning bias will the networks learn given different vocabulary structures. We address this question in a virtual version of the NNG task. On each test trial of the virtual NNG task, we presented the network with three novel entities (one at a time) on the perception layer – an exemplar, and two choice items, one matching the exemplar in shape only and one matching in material only. For each of these three inputs, we recorded the resulting pattern of activation on the hidden layer. This is a measure of how the network represents these items. If the network emphasizes the shape of the item then the similarities of the internal representations for the exemplar and its shape matching choice should be *greater* than the similarity of the internal representations for the exemplar and the material matching choice. If, however, the internal representations highlight the material of the items, then the similarity of the internal representations for the exemplar and the shape matching choice should be *less* than the corresponding similarity of the exemplar and the material matching choice. We used these similarities along with Luce's choice rule to calculate probability of choice using these similarity measures in order to predict performance in the novel noun generalization task.

In previous work these models have been used to demonstrate the plausibility of the idea that the correlations in the early noun lexicon are sufficient to create second order generalizations – knowledge that any solid thing should be named by shape, and any non-solid thing should be named by material. The present simulations extend this work to variable vocabularies of individual children in the bottom and top ends of the language endowment spectrum.

## Results

The networks' predictions for each of the fifteen vocabularies of early talkers and late talkers are shown in Figure 2. In short, using a cut-off of at least two standard deviations above or below the 50% chance level mark, all networks in the early talker group show a shape bias for solids, and 12/15 early talker networks show a material bias for non-solids as well. In contrast, 12/15 late talker networks show a shape bias for solids and only 3/15 show a material bias for non-solids. Interestingly, 6/15 late-talker networks show a shape bias for *non-solids,* a novel prediction that has not been empirically tested so far. To further analyze the networks' performance, networks were classified according to the observed generalization patterns: *correct* if they showed a shape bias for solids and a material bias for non-solids, *half-right* if they showed the appropriate shape bias for solids but no consistent bias for material, or *wrong*, if they showed an incorrect overgeneralized shape biased to non-solids. A chi-square test showed these types of word learning biases were distributed differently in late talker and early talker networks, $X^2(2,15)=14.743$, $p=.0006$ (Yates' $p=0.004$).

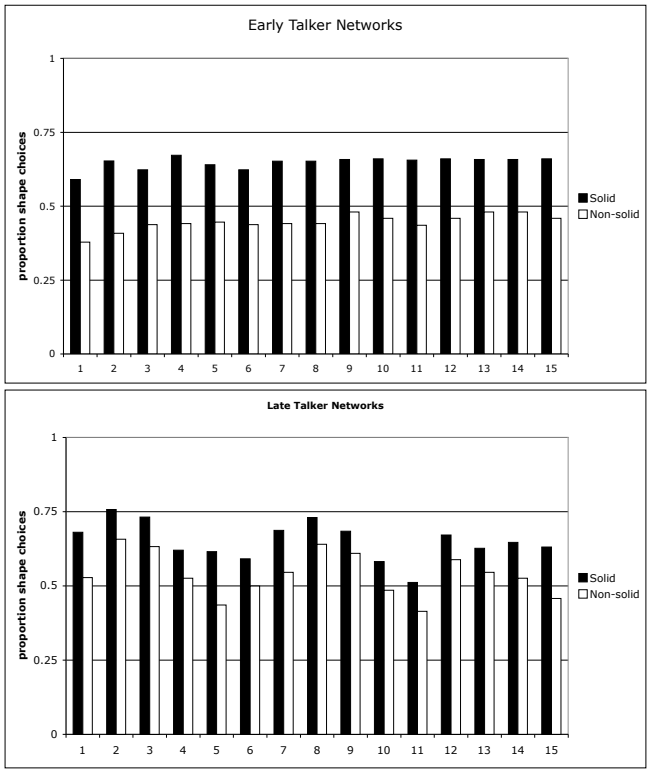


**Figure 2. Predicted proportion of shape choices for each of the early talker and late talker networks**

## Discussion.

The results of the simulations suggest that the differences in noun vocabulary composition between late- and early-talking children may result in differences in word learning biases. The word learning biases learned by these networks can be interpreted as predictions at the group level. First, the networks make a novel prediction about early talkers. A majority of the early talker networks show material biases for non-solids. Previous findings have shown that children at this age (18- to 30-month-olds) show a material bias for non-solids only when offered extra cues. For example, Soja (1992) showed older 2-year-olds have a material bias when offered supporting syntactic and visual cues, and Colunga & Smith (2005) showed an early material bias for non-solids that were presented in simple shapes for older 1-year-olds. However, children in general do not show a robust material bias for non-solids until around age 3 (Samuelson & Smith, 1999). Thus, this is a novel prediction that warrants testing: the networks predict that early talkers, unlike the general population, will show an early material bias for non-solids even without supporting cues.

The networks also make predictions about the patterns of novel noun generalizations one should expect to see in late talkers between 18 and 30 months of age. As a group, late talkers should show a shape bias for solids, with about half of them overgeneralizing this shape bias to non-solids as well. In Experiment 2 we test these predictions with late- and early-talker toddlers in the lab. Additionally, we run neural network simulations based on the composition of the individual vocabularies of these children to replicate the pattern found in Experiment 1.

# Experiment 2

## Method

**Participants**. Nine late talkers (5 girls) and 8 early talkers (4 girls) between the ages of 18 and 22 months (M=19.4) were selected out of 32 children recruited as part of a larger study. As in Experiment 1, the criterion for inclusion was scoring at or below the 25th percentile for late talkers and at or above the 75th percentile for early talkers. MCDI scores ranged from 5th to 20th percentile (M=8.9) for the late talkers and between 75th and 99th percentile for early talkers (M=91). Vocabulary sizes for the late-talker group ranged between 9 and 82 words (*M*=33) and between 151 and 526 words (*M*=376.3) for the early-talker group.

**Stimuli.** The stimuli consisted of a warm up set, a solid set and a non-solid set. The warm up set had an exemplar, a red plastic ball, two other balls (a tennis ball and a green and blue rubber ball), a plastic spoon, a toy carrot, and a toy cat.

The solid set consisted of an exemplar, an orange fuzzy round container, and 5 test items: two items matching the exemplar in shape alone (iridescent green bumpy round container and golden glittery round container), two items matching the exemplar in material (fuzzy blue irregular ring and fuzzy orange hook-like shape, and one matching in color (orange mesh polyhedron). The non-solid set was similarly structured and consisted of an exemplar (purple craft sand mixed into Noxzema in a rounded elongated x-like shape), two material matches (green sand + Noxzema in an asymmetric s-like shape and red sand + Noxzema in a lollypop-like shape), two shape matches (elongated x-like shapes made out of sawdust or purple shaving cream), and a color match (purple hair gel in an hourglass shape. All non-solids were presented on flat, square, plastic foam boards.

**Procedure.** In the warm-up phase, the experimenter presented all six toys to the child and allowed him or her to look at them and handle them for 30 seconds. Then the objects were removed and immediately placed back on the table outside of the child's reach. The child was then shown the exemplar ball and told, "look at this ball." Then they were asked to "get a ball" or get "another ball." If the child failed to retrieve a ball, the child was asked one more time, and finally was told "here's another ball," handed the ball, and allowed to get it one more time on request. If the child got one of the non-ball distracter items, they were told, "that's not a ball, that's a ____" , then the distracter was replaced on the table, and the child was asked again for it.

The procedure during the test phase with the solid and non-solid novel sets was the same, except that no feedback was offered. Children were shown the exemplar and told, "Look at this dax" and then asked to "get a dax" or "get another dax" for the solid set or "get more dax" or "get some dax" in the non-solid set. Children were asked to get another (or more) until they indicated that there were no more. Thus, solids were presented with count syntax supporting an object construal and non-solids were

presented with mass syntax supporting a substance construal (Soja, 1992). The solid set was always presented before the non-solid set, and there was a 5-minute break and a change in testing rooms in between the two test sets.

**Coding.** To incorporate order information into children's choices, and because all children made at least three choices for each test set, their choices were coded as follows: 3 points for the item that was 1st choice, 2 points for the 2nd choice, 1 point for the 3rd choice, and 0 points for other.

## Results

**Simulations.** The simulations based on individual children's vocabularies showed patterns comparable to the ones in Experiment 1. For the early talker networks, 6/8 showed shape and material biases, and the other two showed only a shape bias and no robust material bias. None of the early talker networks showed incorrect biases. For the late talker networks, all eight networks showed a shape bias for solids, but only one showed a material bias for non-solids[1]. In addition, 4/8 late talker networks showed an overgeneralized shape bias for non-solids. A chi-square test showed these types of word learning biases were distributed differently in late talker and early talker networks, $X^2(2,8)=7.77$, $p=.02$ (Yates' $X^2= 4.54$, $p=0.103$).

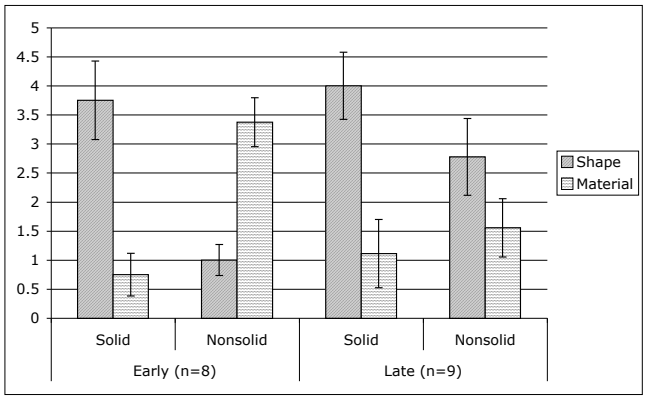


**Figure 3. Scores for shape and material matches for solids and nonsolids for early- and late-talking toddlers.**

**Word learning biases.** The simulations in Experiment 1 predicted that early and late talkers would show different word learning biases, and predict specific patterns of novel noun generalizations for solids and non-solids for these two groups of children. We first look at the data of all children together and then evaluate the predictions for each language group. We submitted both groups of children's scores for the shape and material test items for the solids and non-solid sets to a 2 (language group: early, late) x 2 (solidity: solid, nonsolid) x 2 (dimension: shape, material) mixed ANOVA. Figure 3 shows the average score for the items that matched the exemplar in shape or material for the solid and nonsolid sets for both language groups. There was a main effect of dimension, $F(1,29) = 4.77$, $p = .045$, $\eta^2 = .24$; overall shape matches had higher scores than material matches. There was also a significant interaction between solidity and dimension, $F(1,15) = 15.6$, $p=.001$, $\eta^2 = .51$. Post-hoc tests showed that across both language groups, children were more likely to choose the shape over the material match for the solid set, $t(16) = 4.03$, $p=.001$, but not for the nonsolid set, $t(16) = -.613$, n.s. The three-way interaction between language group, solidity, and dimension was marginally significant, $F(1,15) = 4.33$, $p=.055$, $\eta^2 = .22$.

The language-group-specific predictions made by the models were tested by analyzing the two groups separately. First, the prediction that early talkers would show a robust shape bias for solids and a robust material bias for nonsolids was confirmed by a 2 (solidity) x 2 (dimension) ANOVA revealing a two-way interaction between solidity and dimension, $F(1,7) = 26.15$, $p = .001$, $\eta^2 = .78$. Furthermore, planned comparisons (all two-tailed) showed that this interaction came from early talkers' shape bias for solids ($t(7)=3.06$, $p=.018$) and material bias for non-solids ($t(7)=-4.46$, $p=.003$). Second, a similar analysis on late talkers' scores revealed a main effect of dimension, $F(1,8) = 5.5$, $p=.047$, $\eta^2 = .41$, and no other main effects or interactions. Planned comparisons showed that late talkers had a shape bias for solids, $t(8) = 2.57$, $p=.033$, but did not overgeneralize the shape bias to non-solids as a group, $t(8) = 1.1$, n.s. However, 4 out of the 9 late talkers in the study showed a shape bias for non-solids (a difference score of more than 3), and none of the early talkers did.

## Discussion

The results of Experiment 2 confirm the predictions of the simulations in Experiment 1. Early talkers show a shape bias for solids and a material bias for non-solids; late talkers show a shape bias for solids that can be over-generalized to non-solids. It is important to note that these predictions work at the group level and not at the level of individuals. For example, although four late takers showed an overgeneralized shape bias for nonsolids in both the behavioral tasks and in the network simulations, these were not the same children; only two children showed this bias in both the simulations based on their vocabularies and their performance in the behavioral task. The behavioral task, and probably the vocabulary measure as well, lack the finesse to make predictions at the individual level based on a single data point. We return to this point in the general discussion.

The results of experiment 2 are in line with previous work noting a relationship between the number of nouns in a child's vocabulary and their word learning biases, but they extend it in important ways. The finding that early talkers show robust word learning biases for both solids and non-solids at not even two years of age is new. Although one might have predicted this pattern of results a priori from either the empiricist or the rationalist sides of the word learning debate, or even just from the idea that early talkers might excel across tasks, the prediction came from the models. Harder to predict without the networks, however, is the pattern found for the late talkers. In fact, at first glance it seems to contradict what we know about late talkers; that 2- to 3-year-old late talkers lack a shape bias while their same-

[1] One late talker child had no nouns, so no network was ran for that child. Thus, only 8 late talker networks were ran.

aged peers already have a well-established bias. However, the prediction from the networks, and our findings on the patterns of word learning biases in very young late talkers, before the age of 2, can help us understand the processes underlying word learning in general.

Gershkoff-Stowe and Smith (2004) followed eight children as they learned their first 100 nouns, looking at their word learning biases for solids and their vocabulary growth every three weeks. Their results show that as children's noun vocabulary increases, so does their attention to shape. They set the emergence of the shape bias at around the time children acquire 50 nouns. Our results suggest that this relationship may be different for late talkers. None of the late talkers in Experiment 2 reached the 50-noun mark (though a couple were on the cusp), and yet they overall showed a robust preference for shape for the solid set in our task. Curiously, although attention to shape increased with vocabulary size in Gershkoff-Stowe's study, the lower vocabulary group did show a preference of shape over material. This suggests an intriguing possibility: These models do not make a distinction between naming and non-naming contexts. It is possible that the shape preference for solids here is not a true shape bias, but rather an overgeneralized heightened attention to shape. The fact that about half of the late talkers showed an overgeneralized shape bias for non-solids suggests that this may be the case.

## General Discussion

The work presented here makes two main contributions. First, the findings of these two studies show that late talkers and early talkers know different sorts of nouns that lead to qualitatively different word learning biases. Importantly, these differences are shown within a computational model that has been previously shown to capture various aspects of novel noun learning, suggesting a promising use for process-level computational models. Efforts to tease apart the contributions of different factors to outcomes in late talkers have come up with some characteristics that put children at higher risk, but the underlying mechanisms are not well understood. The work of Ziegler and colleagues in the domain of dyslexia offers a good example of the potential for using computational models – and specifically models that operate at the mechanistic level – in simulating individual differences and further understanding subtypes in atypical development (Ziegler, Castel, Pech-Georgel, George, Aario, & Perry, 2008). Thus, the models presented here are a promising first step in leveraging computational models to aid in the understanding of why some late talkers catch up and others do not.

Second, these models represent an important extension over previous word-learning modeling efforts in that they go beyond modeling the performance of the mythical average child to making predictions about the performance of individual children, and of children who are both at the top and at the bottom of the vocabulary spectrum. In so doing, the simulations presented here make novel and testable predictions. The relationship between vocabulary composition and word learning biases modeled here – the words you know determine the way you learn new words, which constrains and facilitates the words you will know next, and so on – opens a new way of thinking about computational models, to capture not only averages and not only individuals, but individual *trajectories*. If we can build computational models that can successfully capture this self-constructing developmental loop, the implications for early diagnosis, designing early interventions, and understanding the mechanisms that underlie word learning in typical and atypical development are far-reaching.

## References


Colunga, E., & Smith, L. B. (2005). From the lexicon to expectations about kinds: a role for associative learning. *Psychological Review*, *112*(2), 347–382.

Colunga, E. & Sims, C. (2011). Early Talkers and Late Talkers Know Nouns that License Different Word Learning Biases, in *Proc. of Cog. Sci Society, 32*. pp. 2250-2255.

Desmarais, C. S., Meyer, F., Bairati, I., & Rouleau, N. (2008) Systematic review of the literature on characteristics of late-talking toddlers. Intl. J. of Language and Communication Disorders. 43(4) pp:361-389.

Fenson, L., Dale, P., Reznick, J. S., Thal, D., Bates, E., Hartung, J., Pethick, S., & Reilly, J. (1993). *The MacArthur Communicative Development Inventories: User's guide and technical manual*. San Diego, CA:

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

Gershkoff-Stowe, L., & Smith, L.B.(2004). Shape and the first hundred nouns. *Child development 75*(4),1098-1114.

Jones, S. S. (2003). Late talkers show no shape bias in a novel name extension task. *Dev. Science, 6*(5), 477-483.

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

Rescorla, L. (2002). Language and reading outcomes to age 9 in late-talking toddlers. *Journal of Speech, Language, and Hearing Research,* 45, 360-371.

Rescorla, L., Roberts, J. & Dahlsgaard, K. (1997). Late talkers at 2: Outcome at age 3. *Journal of Speech & Hearing Research,* 40, 556-566.

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

Smith, L. B., Jones, S. S., Landau, B., Gershkoff-Stowe, L., & Samuelson, L. . (2002). Creating a shape bias creates rapid word learners. *Psychological Science, 13*, 13-19.

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

Ziegler, J. , Castel, C., Pech-Georgel, C., George, F., Alario, F-X., & Perry, C. (2008). Developmental dyslexia and the dual route model of reading: Simulating individual differences and subtypes. *Cognition, 107*, 151–178.