Spring 2014

Modeling Developmental Word Learning in Late Talking Children

Katherine Double

University of Colorado Boulder

Follow this and additional works at: http://scholar.colorado.edu/honr_theses

Recommended Citation


This Thesis is brought to you for free and open access by Honors Program at CU Scholar. It has been accepted for inclusion in Undergraduate Honors Theses by an authorized administrator of CU Scholar. For more information, please contact cuscholaradmin@colorado.edu.
Modeling Developmental Word Learning in Late Talking Children

Katherine Double

Honors Thesis, Spring 2014

Defense: March 14, 2014

Committee Members:

Primary Advisor: Eliana Colunga, Assistant Professor, Department of Psychology

Richard Olson, Honors Council Representative, Department of Psychology

Patricia Mcshane, Instructor, Special Undergraduate Enrichment Programs
Abstract

In typical early vocabulary development, young toddlers are skilled at learning noun categories. In fact, they use attentional word learning biases in order to categorize nouns, which helps increase their vocabulary. In particular, in a Novel Noun Generalization task, they show a shape bias for solid objects, and a material bias for non-solid substances. Children who are at a delayed vocabulary level compared to their peers, or late talkers, do not exhibit the same word learning biases. A sample of 33 late talkers showed that those who remain persistently delayed in the long term differ from other late talkers in both their noun vocabulary structures and their word learning biases. Behavioral data using the Novel Noun Generalization task shows that the group of long term persistent late talkers developed a shape bias that was not significantly larger than the material bias, whereas the other two groups of late talkers had a significantly larger shape bias than material bias. Computational models that simulate the Novel Noun Generalization task show that the long term persistent late talkers develop a shape bias and a material bias faster than the other two groups. This tell us that this particular group of late talkers stands out from the others, and may not develop biases in the same way, which could be the cause of their vocabulary delay. The direction of this difference is not entirely clear, as the models do not replicate the behavioral data. However, acknowledging a fundamental difference in the way different types of late talkers learn new words is a step toward creating early interventions, and further replications of these studies could help the network simulations further match behavioral data seen in the lab.
Introduction

Children are amazing word learners, although they vary greatly in their language acquisition abilities. While many children show no struggle in learning new words and expanding their vocabulary, some groups of children have been identified as being at a delayed vocabulary level compared to their peers. Late talkers, as they are commonly known, are particular subjects of interest, both in our studies and in past research. Many late talkers end up catching up to their peers—others do not. It is not always clear why some of the late talkers are able to catch up, and what causes others to remain delayed, as they are not a homogenous group, but rather vary greatly in their clinical, personal, and social outcomes (Desmarais, Meyer, Bairati & Rouleau, 2008). In addition, various factors that contribute to a child’s vocabulary development, such as underlying clinical problems, a social or familial environment that does not help promote vocabulary growth (Carson et al. 1998, Irwin et al. 2002), a child’s personal characteristics, or cognitive processes more directly involved in word acquisition, such as lexical acquisition, communicative intent, use of communicative gestures, and phonetic and phonological skills (Desmarais et al., 2008). Therefore, the origin, course, and outcome of a late talker’s vocabulary abilities must be looked at more on an individual basis, rather than as a group. Although we recognize the complex and dynamic relationship of all of these factors in late talking toddlers, we will focus on their vocabulary structure and word learning attentional biases, rather than the clinical or social factors that play a role in late talkers’ delayed vocabulary growth.
Early word learning in typically developing children

When learning new words, young children show patterns in the way they learn new words, in particular, nouns. By the age of two, typically developing children seem to know a lot about how object categories are named. Certain features, such as solidity (whether it is a solid object or a non-solid substance) and both the shape and material of the object, become important in distinguishing categories named by nouns. During their first few years of language development, children begin to exhibit word learning biases. The task used to study children’s assumptions about the relevant features for different categories is called a Novel Noun Generalization task. Specifically, in this task, a child is shown a novel target object and told its name. The child is then asked which other objects, matching the target in different features, also share that name. Thus, the child has to generalize that label to other category members. In this task, typically developing children show a shape bias for solid objects, which refers to a child’s tendency to generalize novel nouns to other solid objects that are similar in shape (Landau, Smith & Jones, 1988). Multiple studies are consistent in their findings that children start showing robust word learning biases around 2-3 years (e.g., Colunga & Smith, 2005; Yoshida & Smith, 2003; Jones, Smith & Landau, 1991). Children also show a material bias for non-solid substances, in learning to extend a novel name for a non-solid substance to other substances that match in material (Soja, Carey, & Spelke, 1991). Although both these biases are strong, the material bias is seen later than the shape bias, typically around 3 years of age (Yoshida & Smith, 2005), whereas the shape bias emerges during the second year of life. The development of these biases may be a critical step in word learning, as it allows children to use a variety of information in deciding what new words refer to, using features of certain objects to distinguish between them, and learning which features are the most useful for naming different kinds of categories. It
is therefore possible that a deficit in the development of these biases can delay a child’s vocabulary growth.

There is evidence that a relationship exists between a child’s vocabulary composition and the development of word learning biases (Colunga & Smith, 2005; Samuelson, 2002; Samuelson & Smith, 1999). By the time children have between 50 to 150 nouns in their vocabulary, their tendency to attend to shape when naming solid objects emerges particularly strongly (Gershkoff-Stowe & Smith, 2004). Changing young children’s vocabulary composition through intensive teaching of names for shape-based object categories allows them to develop an early shape bias, which in turn accelerates their learning of other objects and increases their vocabulary (Smith, Jones, Landau, Gershkoff-Stowe & Samuelson, 2002). This suggests the presence of a developmental feedback loop, exhibiting a development of biases through the growth of vocabulary, which in turn helps grow the vocabulary even more.

Identifying late talkers

In the studies reviewed above, a child’s vocabulary is evaluated using a variety of measures of productive vocabulary size. Our particular measure of interest in the current work is a standardized vocabulary checklist known as the MacArthur-Bates Communicative Developmental Inventory, or MCDI, which is given to the parents. The parent marks which words on the checklist the child says; the results are then used to place the child in a percentile of productive vocabulary knowledge based on norming data from a large national sample of children (Fenson, Dale, Reznick, Thal, Bates, Hartung et al., 1993). “Late talkers” is the term applied to those below the 25th percentile in productive vocabulary, although there is variation in this arbitrary cut-off percentile score used to categorize late talkers. The MCDI is effective at identifying late talkers, and has empirically proven validity when correlated with direct language
measures (Heilmann, Weismer, Evans & Hollar, 2005). Heilmann and colleagues further remark that the MCDI is most effective in identifying low and high levels of language performance, and less effective at intermediate classification. That is, children who score below the 11th percentile in the MCDI are very likely to show persistent language impairments, and children who score above the 49th percentile are likely to develop language skills normally; however, children between the 11th and 49th percentile cannot be classified. By combining vocabulary, word learning, and neural network simulations, the current work attempts to distinguish those late talkers who will catch up from those who will not.

Although some late talkers remain behind for years, most late talking children eventually catch up to their peers and have no lasting language impairments. However, there is a notable difference in their ability to form attentional biases. Late talkers aged 25-41 months do not systematically generalize the name of a novel solid object to other objects that match in shape the way typically developing children do (Jones, 2003). In fact, in Jones’s study, about half of the late talkers showed the wrong attentional bias, and instead extended the names of solid objects to objects of the same material. A material bias in naming solid objects has not been observed in previous experiments on attentional biases. It is possible that using an incorrect bias in attempting to categorize words and learn new labels makes it more difficult for these children to expand their vocabulary. This could potentially be the mechanism behind the late talkers’ delayed vocabulary growth.

**Network simulations**

The computational models that simulate word learning in children have been used in multiple studies to investigate the development of word learning biases and the conditions that support them. Colunga and Smith (2005) created a neural network that learned a vocabulary
structured like that of the average 30-month-old, based on the MCDI norms. Figure 1 shows the vocabulary structure of a typically developing child according to its distribution of nouns that refer to solid or non-solids in categories that are organized by similarity in shape, material, or both. After learning a vocabulary with that structure, the neural network showed a shape bias for solids and a material bias for non-solids, using a virtual simulation of a novel noun generalization task seen above. Colunga and Sims (2011) used a modified version of this network to model differences between late and early talkers.

<table>
<thead>
<tr>
<th></th>
<th>Shape</th>
<th>Material</th>
<th>Both</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Solid</strong></td>
<td>52%</td>
<td>10%</td>
<td>12%</td>
</tr>
<tr>
<td><strong>Nonsolid</strong></td>
<td>4%</td>
<td>16%</td>
<td>6%</td>
</tr>
</tbody>
</table>

*Figure 1. The structure of a typically developing child’s early vocabulary. The vocabulary is primarily comprised of solid words categorized by shape (such as ball), and includes fewer non-solid substances categorized by material.*

Late and early talkers also show a difference in their noun vocabulary structure (Colunga & Sims, 2011). Specifically, Colunga and Sims looked at the vocabulary structure of 15 early talkers and 15 late talkers matched by age. Aside from differences in vocabulary structure, such as a greater proportion of nouns for solid shape-based categories in late talkers compared to early talkers, late talkers generally tend to have more variability in their vocabulary structure, maybe in part because they know fewer words. In both computational models and the sample of late and
early talking children, differences in word learning biases are also found. Simulations of early talkers showed a shape bias for solids and a material bias for non-solids. Late talkers, however, showed a shape bias for solids, less of a material bias for non-solids, and surprisingly a shape bias for non-solids, once again overgeneralize the shape bias (Colunga & Sims, 2012). This difference is attributed to the difference in vocabulary structure between late and early talkers.

More recent work has looked at the time course of the emergence of these word learning biases in networks and children, using longitudinal data for children and measuring the networks as they learn more and more words. When trained on a typical 30-month-old child’s vocabulary structure, neural networks showed that as the shape bias emerges, other word learning biases diminished (Schilling, Sims & Colunga, 2012). In particular, the emergence of the shape bias was determined as when the network chose the shape match for solids more than 55% of the time. This was found to occur on average when the network learned 15 words out of 100 in the training set. During this emergence, the network’s material preference diminishes over time as the shape preference increases, suggesting that the emergence of the shape bias may slightly influence the material bias.

If there is a feedback loop between vocabulary structure and word learning biases, not only will different vocabulary structures lead to different word learning biases, but acquiring a word learning bias should be associated with a change in learning rate for the sorts of words it favors, and thus a change in vocabulary structure. Sims, Schilling, and Colunga (2013) showed that in networks trained on a typically developing child’s vocabulary, there was an equal increase in the amount of shape- and material-based words in the networks’ vocabularies up to the emergence of the shape bias, after which the learning of shape-based words increased faster than that of material-based words. Combined with results from Schilling et al. (2012) and Colunga
and Smith (2005), this supports the previously discussed idea of a developmental feedback loop: networks are able to learn attentional biases based on a particular vocabulary, which in turn influences subsequent vocabulary learning.

The current study

Overall, it is evident that there is a fundamental difference in the way late talkers learn new words compared to their peers. We now question whether or not there is a difference in both vocabulary structure and the development of word learning biases between late talkers who remain at a delayed vocabulary level, and late talkers who eventually catch up to the productive vocabulary level of their peers and are no longer labeled as late. Using data from two similar year-long longitudinal studies on children 18-30 months old, where late talkers were identified, we can separate those who remain late talkers and those who catch up, and then use network simulations as well as behavioral data from these children in the Novel Noun Generalization task to test whether there is a fundamental difference in the way these two groups of children create categorical biases to learn new words. This could help us understand why some late talkers catch up to their peers and why some continue to have language deficits later in life.

Methods

Participants

Subjects from two nearly identical waves of a year-long longitudinal study were chosen for analysis. The two waves were collected in different years, resulting in two cohorts of children. Late talkers were identified as having an MCDI percentile of 30 or less at the first experimental session. Although the two longitudinal studies were done over a period of twelve months, only the first six months of data were used. This was done to minimize attrition rates. However, the MCDI percentile at the twelfth visit of the participants was noted in order to
categorize the subjects by vocabulary trajectory. Any subjects who missed more than one appointment in the six-month period were excluded. This led to a total of 33 late talkers, 15 male and 18 female (mean age at first visit was 17.5 months, $SD = 1.1$, range = 15.8 to 19.4; mean age at last visit was 22.7 months, $SD = 1.2$, range = 21 to 25.8). Those who had a final MCDI percentile of less than 30 at session six (or session five for those who did not have MCDI data at session six—this only included two subjects) OR at session 12 were labeled as persistent late talkers ($n=15$), and the rest were non-persistent ($n=18$). Upon further analysis, a third group of long-term persistent late talkers was added. This group had a further constraint of needing an MCDI percentile of less than 30 at both the 6th and the 12th visit. This consisted of five of the original 15 persistent late talkers.

**Procedures**

*Behavioral Study*

The subjects were brought in to the lab each month for twelve months to complete three word learning tasks, and the parents filled out the MCDI. The tasks included a warm-up task, a solid novel noun generalization task (NNGT), and a non-solid NNGT. In the warm-up task, the experimenter showed the child six familiar objects (three balls, a spoon, a clip, and a wheel), and then labeled the target object (one of the balls). The experimenter then asked the child if there were any more balls, until the child stopped choosing objects. The child was provided feedback on the warm-up task, but not the subsequent tasks. The goal was to familiarize the child with the task, letting them know that some, but not all of the objects presented, matched the target. In the solid NNGT, novel solid stimuli with novel names were presented to the child (the stimuli set contained a target object; two objects that matched the target in shape but not material or color; one that matched the target in color, but not shape or material; one that matched in material but
not shape or color; and one that matched in both color and material, but not shape—see Figure 2). This created two material matches, two color matches, and two shape matches in the stimuli set. Similar to the warm-up task, the target object was identified and labeled. The experimenter asked the child if there were any more of that object in front of them, and the responses, including the order in which children chose objects, were recorded. The non-solid NNGT was the same as the solid task, but the objects were non-solid substances (this set contained a target substance; two substances that matched the target in material, but not color or shape; one material that was arranged to match the target in shape, but not color or material; one that matched the target in color, but not shape or material; and one that matched the target in both color and shape, but not material).

After the child completed the tasks, the parents filled out the MCDI. They were told to mark all the words the child says (specifically, the words a child uses on his or her own, rather than words a child merely recognizes). It should be noted that the procedure for filling out the MCDI was different for the first and second waves of the study. In the first wave of the study, parents were given a blank MCDI form each time, so it was possible that words that were checked off in one visit were not checked again at the subsequent visit. In the second wave, parents were given a photocopy of their MCDI from the previous visit, so words were only added, and could not disappear from one month to the next. We controlled for this difference between study waves by making the assumption for the first wave that if the word on the MCDI was ever known, it was known on all the subsequent MCDI forms. This was to ensure that the MCDI data from the two waves were consistent.
Network Simulations

The networks were trained on input representing the structure of a particular child’s vocabulary. The raw MCDI data for each subject in their final visit was broken down by the number of nouns learned in six categories, divided by solidity (solid or non-solid) and feature (shape, material, or both). The structure of the vocabulary is expressed as proportions of words in each of the six categories. These proportions were normalized so the total number of word inputs equaled 100. The vocabulary structure then consists of input patterns created individually for each child.

The network consists of a word layer, where word labels are input; a perceptual layer, where features (shape, material, or both) are represented as distributed patterns across the layer and solidity is represented discretely, with one unit that stands for solid and another that stands for non-solid; and the hidden layer, which connects the two other layers and represents where learning occurs (see Figure 3). The hidden layer allows the network to build internal representations to stand between the objects (solid and non-solid) and their labels.
Figure 3. The word layer connects to the perceptual layer through the hidden layer. The perceptual layer represents the features and solidity of the objects.

The network was then trained on each individual child’s vocabulary. During training, the network formed biases based on the vocabulary structure that was used as input. This structure was represented by a pattern of features that was paired with a word for each trial of training. A randomly selected input vector would then represent that word’s features. For instance, to make a solid-shape word (a solid object categorized by shape), a unit in the word layer (representing the name of the category) and the unit representing solid were activated with a 12-bit randomly generated binary number to represent the shape of the object. Since the network will run through the entire vocabulary input multiple times throughout the course of learning, it will see several instances of each kind of word, analogous to the way a child will see multiple instances of a word before deciding what its label means. Therefore, the randomly generated shape part of the word stays the same each time the network sees an instance of the word. The value of the material pattern was randomly assigned, so this pattern was different each time. Patterns for
nonsolid-material words were generated in the same way, except that the shape pattern varies and the material pattern remains the same. This was done for all 100 words in each vocabulary input. The network was trained for 500 epochs, so over training, it was presented with 500 instances for all the words in the training set. Words learned for each category were recorded for multiple time points in the course of learning the entire vocabulary, based on the number of words learned at each visit in the MCDI. For example, the network would stop and record the words learned in each category after it learned 10 words, then 17, then 28, 53, 76, and 100 words. These numbers are the number of words this particular subject knew on the MCDI at each of the 6 visits (normalized to 100 total words). This approach uses children’s individualized vocabulary structures as input, and observes the course of learning based on real children’s individual learning patterns, which differs from past research on networks using the same time points for all networks.

Simulating the NNG tasks done in the lab, the networks were tested for three runs at each of the time points with a solid task and a non-solid task. This was to see if after learning the vocabulary up to a given point, the network attended to the shape of a solid object and the material of a non-solid substance. Each trial of each testing task presented the network with novel word representations including an exemplar, an item matching the exemplar in shape only, and an item matching the exemplar just in material. The resulting hidden layer pattern was then recorded as a prediction of performance on the NNG task. If the network learned to attend to shape in the context of solidity, then it would be expected that given a solid input pattern, the pattern of activation on the hidden layer should represent mostly the input’s shape information rather than material. The opposite is true for material in the context of nonsolidity.
Results

Vocabulary structure

We analyzed the subjects’ vocabulary structures with a 6(Visit) x 2(Solidity: solid, non-solid) x 3(Feature: solid, material, both) x 3(Talker Type: non-persistent, persistent, long-term persistent) repeated measures ANOVA. The outcome measure was the proportion of each word type known (solid-shape, solid-material, solid-both, nonsolid-shape, nonsolid-material, and nonsolid-both). We used proportions rather than raw counts to normalize across vocabularies of different sizes. There was a main effect of solidity, $F(1.24) = 1374.703$, $p<0.001$, and a main effect of feature, $F(2.48) = 99.003$, $p<0.001$. This is what we would expect as a result of using the MCDI, since the six different word categories are represented differently on the checklist. For example, there is a higher proportion of shape-based solid nouns ($M = 0.654$, $SD = 0.172$) than material-based non-solid nouns ($M = 0.045$, $SD = 0.034$). There was also an interaction of visit, solidity, and feature, $F(10.240) = 3.756$, $p<0.001$. The interaction with visit shows the vocabulary structure changes over time (see Table 1), with the proportion of solid-shape words decreasing over time.

Table 1.

Percentages of each word type in vocabulary by visit.

<table>
<thead>
<tr>
<th>Visit</th>
<th>Solid-Shape</th>
<th>Solid-Material</th>
<th>Solid-Both</th>
<th>Nonsolid-Material</th>
<th>Nonsolid-Both</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>74.7</td>
<td>2.5</td>
<td>16.4</td>
<td>5.2</td>
<td>1.3</td>
</tr>
<tr>
<td>2</td>
<td>74.7</td>
<td>3.4</td>
<td>15</td>
<td>3.8</td>
<td>2.7</td>
</tr>
<tr>
<td>3</td>
<td>64.9</td>
<td>4.3</td>
<td>22.1</td>
<td>4.9</td>
<td>3.8</td>
</tr>
<tr>
<td>4</td>
<td>65.9</td>
<td>4.4</td>
<td>21.9</td>
<td>3.9</td>
<td>3.9</td>
</tr>
<tr>
<td>5</td>
<td>64.6</td>
<td>5.2</td>
<td>23.5</td>
<td>3.4</td>
<td>3.3</td>
</tr>
<tr>
<td>6</td>
<td>62.1</td>
<td>6.1</td>
<td>24.4</td>
<td>3.6</td>
<td>3.9</td>
</tr>
</tbody>
</table>

There are no nonsolid-shape words on the MCDI; however, this word type was kept in analysis in order to keep the design balanced, so there were three feature types for each solidity type.
There was a three-way interaction between visit, solidity, and talker type, $F(10,120) = 2.213, p = 0.021$. There was no four-way interaction, suggesting that whereas the three groups differ on their proportion of solid and non-solid words known over time, the feature of the nouns do not play into this effect. We further analyzed these effects by conducting ANOVAs at each visit. These analyses showed there was no effect of talker type in the first three visits. However, visits 4-6 showed a significant interaction between solidity, feature, and talker type. The interaction grew stronger over the last three visits (Visit 4: $F(4,58) = 3.845, p = 0.008$; Visit 5: $F(4,60) = 5.26, p = 0.001$; Visit 6: $F(4,56) = 6.585, p < 0.001$), and visit 6 exhibited the strongest interaction (Figure 4).

![Figure 4. Visit 6 vocabulary structure by talker type.](image)

**Behavioral Data**

The outcome measure for the behavioral study was coded attention scores on the NNGT performed in the lab. Subjects were scored based on their shape choices for the solid set, and material choices for the non-solid set (refer to Figure 2). They would receive a score of 3 for
their first choice, 2 for their second choice, 1 for their third choice, and a 0 for any other subsequent choices. We ran a 6(Visit) x 2(Trial Type: solid-shape, nonsolid-material) x 3(Talker Type) repeated-measures ANOVA. There was a main effect of trial type, $F(1,14) = 10.305, p = 0.006$. The attention score was higher for solid-shape trials ($M = 2.868, SD = 1.27$) than for nonsolid-material trials ($M = 1.786, SD = 1.126$), suggesting that overall, subjects pay more attention to shape on solid trials than to material on non-solid trials. This is consistent with what we would predict in the NNGT in lab: the shape bias is stronger than the material bias. Although there was no significant effect of talker type in the omnibus ANOVA, talker type groups were analyzed with separate 6 (Visit) x 2 (Trial Type) analyses to better understand patterns within groups. When talker types were analyzed separately, they showed different trial type effects. The non-persistent late talkers had a main effect of trial type, $F(1,12) = 20.058, p = 0.001$, as did the persistent late talkers, $F(1,4) = 17.057, p = 0.014$. The long term persistent late talkers, however, did not have a significant difference between trial types ($p = 0.128$). See Figure 5.

![Figure 5](image_url)

**Figure 5.** Attention scores in solid and nonsolid trials by talker type with means.
Network Simulations

The outcome measure for the networks was the proportion of shape choices on the solid test, and the proportion of material choices on the non-solid test. This is analogous to the NNGT in lab. The proportions were ran through a 7(Visit) x 2(Trial Type) x 3(Talker type) repeated-measures ANOVA². There was an effect of visit, \( F(6,180) = 324.346, p < 0.001 \), and trial type, \( F(1,30) = 1078.301, p < 0.001 \), as well as an interaction between visit and trial type, \( F(6,180) = 333.77, p < 0.001 \). The interaction shows that the difference in choice on the virtual NNGT changes over time, something we notice in the behavioral task with children in the lab as well. The proportion of shape choices on solid trials (\( M = 0.706, SD = 0.034 \)) was higher than the proportion of material choices in non-solid trials (\( M = 0.521, SD = 0.006 \)), showing that the direction of the trial type effect is also the same as what we observe in the lab.

We also looked at differences between each talker type by conducting ANOVAs for solid and non-solid trials separately, as seen in Table 2. In solid trials, the long term persistent group mean was significantly higher than both the persistent, \( F(1,13) = 16.356, p = 0.001 \), and non-persistent groups, \( F(1,21) = 21.57, p < 0.001 \). The difference between the non-persistent and persistent groups was not significant. In non-solid trials, the long term persistent group mean was significantly higher than the persistent group, \( F(1,13) = 6.132, p = 0.028 \), but there were no significant differences between the long term persistent and non-persistent groups, or between the persistent and non-persistent groups.

---

² It should be noted that the network has seven rather than six “visits”, or time points, because it is initialized before training and testing, and the initial weights were recorded and included as the initial time point.
Table 2.

Proportion of choice means in solid and nonsolid trials by talker type.

<table>
<thead>
<tr>
<th>Trial Type</th>
<th>Talker Type</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Solid-shape</td>
<td>Non-persistent</td>
<td>0.681</td>
<td>0.030</td>
</tr>
<tr>
<td></td>
<td>Persistent</td>
<td>0.683</td>
<td>0.032</td>
</tr>
<tr>
<td></td>
<td>Long term persistent</td>
<td>0.754</td>
<td>0.031</td>
</tr>
<tr>
<td>Nonsolid-Material</td>
<td>Non-persistent</td>
<td>0.520</td>
<td>0.008</td>
</tr>
<tr>
<td></td>
<td>Persistent</td>
<td>0.517</td>
<td>0.006</td>
</tr>
<tr>
<td></td>
<td>Long term persistent</td>
<td>0.526</td>
<td>0.007</td>
</tr>
</tbody>
</table>

There was also a main effect of talker type in the omnibus ANOVA, $F(2,30) = 11.565$, $p < 0.001$, as well as a significant three-way interaction between visit, trial type, and talker type $F(12,180) = 4.016$, $p < 0.001$. This shows that the talker types differed in their proportion of choice on both solid-shape and nonsolid-material trials over time. We then separated the trial types to analyze the effect of talker type for each, and found an interaction between visit and talker type for the solid-shape trials $F(6,180) = 349.724$, $p < 0.001$, as well as for the nonsolid-material trials, $F(6,180) = 21.896$, $p < 0.001$. Figure 6 shows that the long term persistent group diverges from the other two groups in the solid-shape trials over time, whereas the non-persistent and persistent groups remain relatively similar.

![Figure 6. Proportion of word type choice over time.](image-url)
Discussion

In this study, I analyzed the differences in both vocabulary structure and attentional biases among persistent, non-persistent, and long term persistent late talkers. It is evident that the vocabulary structures are different among the groups. The fact that this effect is significant in visits 4-6, and strongest by visit 6, shows that we can tell the difference among late talkers not only in vocabulary size (the measure we used to separate them in the first place), but vocabulary structure as well. By visit six, long-term persistent late talkers have the highest proportion of solid-shape words in their vocabulary, but the lowest proportion of other words. In the networks, the vocabulary structures of this group also led to the development of a shape bias the fastest out of all the groups, although the behavioral data does not reflect this. In the behavioral data, both non-persistent and persistent late talkers show increased attention to shape in solids relative to attention to material for non-solids; this pattern suggests a stronger shape bias than a material bias. However, the long term persistent late talkers do not show a significant difference between these two biases, though they showed a trend of a smaller shape bias and larger material bias over the other two groups. The fact that the behavioral NNGT and the networks’ simulated NNGT diverge for long term persistent late talkers suggests that these models do not capture their learning well.

The network data suggests that long-term persistent late talkers are paying too much attention to shape, and not enough to other features that may be important. For instance, they could have difficulty extending names of solid objects to those similar in both shape and material, since they may ignore features that are not shape. Past studies (Schilling, Sims & Colunga, 2012; Jones, 2003) show a tendency for late talkers to overgeneralize a shape bias, so this may be a characteristic that is enhanced in long-term persistent late talkers. However, the
behavioral data suggests that although the long term persistent late talkers are still forming biases, they do not show the typical pattern of development for the two biases. Specifically, they do not have a significantly larger shape bias over the material bias. Rather than suggesting an overgeneralization of a shape bias, this suggests the opposite: long term persistent late talkers are not paying enough attention to shape, although they are paying more attention to material when they should be. This may help explain the apparent discrepancy between Jones’ (2003) work showing 25-41 month old late talkers lacked a shape bias, and Sims & Colunga’s (2012) work showing 18-30 month old late talkers had an overgeneralized shape bias—the late talkers tested by Jones were long term persistent by our current definition. The evidence also suggests the possibility of a trade-off: long term persistent late talkers are more flexible in their ability to pay attention to the right feature when they should, but this may come at the cost of less attention to shape for solids compared to the other late talkers. Because the shape bias helps children accelerate their word learning, not paying enough attention to shape could potentially be a cause of this particular group’s delay in vocabulary.

In the vocabulary structure, behavioral data, and the network simulation outcomes, the persistent and non-persistent late talkers are relatively similar, while the long-term persistent late talkers were different from both groups. This suggests that although the persistent late talkers do remain delayed for a period of time, they may perhaps be expected to eventually catch up just as the non-persistent late talkers do. These two groups, although originally labeled as late talkers, may end up having more in common with average or early talkers than they do with what we traditionally think as late talkers. In fact, some of the non-persistent late talkers end up above the 80th percentile a year after we identified them as late talkers. They may have had a brief delay in talking, but recovered quickly. This is perhaps what sets the persistent and non-persistent late
talkers apart from the long-term persistent group. Although the non-persistent late talkers had the vocabulary level of late talkers at one point, they may not have had other characteristics typical of late talkers, whereas the long-term persistent late talkers have certain traits that make it more difficult for them to learn new words, something that makes them inherently late talkers. This is expected, considering that late talkers as a whole are a non-homogenous group.

Another consideration to be made is that the MCDI is a measure of productive vocabulary. Children often times know more words than they say. This is a potential shortcoming to our studies, as some late talkers may have a lot of words in their vocabulary, and may have even learned them using correct attentional biases, but there may be some other factors causing them not to say these words. This could also help explain why we see the jump in percentile score for those who catch up. Perhaps they always knew a lot of words, but suddenly started to verbally communicate more. Again, they might not have really been “late talkers” in the traditional sense, but rather average or early talkers who either found a way to catch up or were delayed only in expressing the words that they already knew.

So what allowed the original late talkers to catch up? If the network simulations are accurately predicting performance on the NNGT by actual children, there is a relationship between the words that children know and the way they learn new words. Having different vocabulary structures and sizes predicts different performance outcomes on the NNGT, and therefore different development of word learning biases. However, the direction of this effect is not what we would expect in the networks. The long-term persistent late talkers are developing a shape bias faster. Colunga and Sims (2012) showed that late talkers had a shape bias for solids, but overgeneralized this to non-solids. This was discussed as a potential reason for their delay in vocabulary. Considering these results, it may be tempting to describe the long-term persistent
late talkers in the same way—however, it does not appear that the long-term persistent networks are learning a shape bias at the expense of a material bias, or overgeneralizing the shape bias. They are actually developing a material bias faster than the other two groups as well. It is possible that there is another mechanism behind the delay in the long term persistent networks’ vocabulary.

It is important to recognize the implications of using the networks as models that may not perfectly predict actual performance of children. The behavioral data suggests something different than the network data does. The failure of the networks to account for the behavioral data indicates that further adjustments to the model will be necessary in order to account for the full spectrum of developmental trajectories. Since they are only models, they need to constantly be tested and revised. Although the networks have successfully matched the behavioral data in past research, this does not imply that we have created ideal models of learning, but are getting closer in making accurate predictions of how children will behave in the lab. The studies performed in lab are also not perfect—there is always a possibility for errors when conducting tasks with an age group like toddlers. Many times the toddlers are not compliant with the tasks, or they may not understand the task but still choose objects at random. The difference in outcomes between the behavioral and network data can partially be explained by the inaccuracies and the inherent noise in both methods.

In dividing late talkers by their persistency, we can learn new ways to help the group of late talkers as a whole catch up to their peers, as well as analyze differences between the types of late talkers and use this information to help truly persistent late talkers in different ways than late talkers that have a higher potential to catch up. It is important to analyze vocabulary structure after a specific point in time, as the difference may not be detectable until children reach a
certain age. The vocabulary structures did not differ between talker types until partway through the study. It is still not entirely clear why these long-term persistent late talkers are still delayed when other children who were originally late talkers as well had an easier time catching up, but we may know now to treat them as a separate group altogether. Perhaps they are developing a shape and material bias better than other late talkers, as the network data may suggest, or maybe there is a trade-off in attention to material at the expense of decreased attention to shape, as the behavioral data suggests. In the former case, there may be a completely different mechanism behind the delay in their vocabulary, something to be further identified and studied. In the latter case, helping them accelerate the shape bias could greatly benefit their vocabulary development. Both cases support the future production of early intervention, and can assist late talkers in catching up to their peers.

Limitations

A limitation in using the MCDI to label late talkers is the somewhat arbitrary categorization of groups. Using a 30th percentile cutoff may reduce the specificity of the test, as it may be capturing children who are not really late talkers. As mentioned before, many late talkers catch up, so that leaves only the late talkers that are truly persistent in the long term, which in our case, is a small sample size of five. A possible cause for having so few children who stay late talkers may be a non-representative sample of the population. The children in our studies are recruited from the Boulder Country area, and may therefore have on average a higher socioeconomic status, which may predict better vocabulary skills if they have a better access to resources. This may cause us to observe higher average MCDI percentiles than are represented in the general population. Perhaps using a higher percentile cutoff is then helpful in setting aside children who are delayed compared to their immediate peers, who have higher than average
percentile scores, rather than those who are delayed compared to the average population. Still, decreasing specificity in order to increase sample size may be adding data that does not help compare true late talkers.

**Future Directions**

There are different waves of this study, and it is ongoing. Therefore, the next step would be to continue testing more participants in more waves of the longitudinal studies. This can help increase the number of late talkers in the pool of subjects, and allow for more accurate results. We also have access to follow-up data collected approximately six months, and a year after each longitudinal study ended. Adding this data would show us an even longer trajectory of the late talkers, and create a longer measure of persistency. By the follow-up visits, even fewer late talkers remain, as some more catch up during the year between those visits. Once again, conducting more waves of the study would increase the sample size of persistent late talkers, especially in follow-up studies. Another step could be running more networks that include some trained on early talker vocabulary structures to use as comparison. This can identify if the differences between non-persistent late talkers and early talkers are larger or smaller than the difference between these late talkers and persistent late talkers. Additionally, this data can be applied in the creation of interventions for late talkers. Teaching specific types of words or further training on different biases can possibly fix the disruption in the developmental feedback loop that makes it more difficult for children to simultaneously learn new words and form biases.

The failure of the networks to replicate the behavioral data leaves room for improvement in the networks. Perhaps if we can adjust the networks to more closely fit the data we see in the lab, they can be a more useful tool in predicting outcomes in an NNGT. The networks still picked up on group differences, though they were not in the direction we would expect. Another
step we could take in the future is collecting early vocabulary data from kids already known to have language deficits. Here, we would be retroactively examining late talkers’ vocabulary development to find an underlying possible cause for their delay. Once the networks are revised, they could be used as a diagnostic tool using this early vocabulary data.
References


