Taking Development Seriously: Modeling the Interactions in the Emergence of Different Word Learning Biases

Savannah M. Schilling (Savannah.Schilling@Colorado.Edu)
Department of Electrical, Computer, & Energy Engineering, 425 UCB
Boulder, CO 80309-0425 USA

Clare E. Sims (Clare.Holtpatrick@Colorado.Edu)
Department of Psychology & Neuroscience, 345 UCB
Boulder, CO 80309-0345 USA

Eliana Colunga (Eliana.Colunga@Colorado.Edu)
Department of Psychology & Neuroscience, 345 UCB
Boulder, CO 80309-0345 USA

Abstract

Development is about change over time. Computational models have provided insights into the developmental changes seen in different cognitive phenomena, including within the domain of word learning. The present paper uses a computational model to investigate the interdependencies between the emergence of different word learning biases. This model allows investigation of how the emergence of the shape bias influences novel noun generalization to two other types of items. The results suggest that the emerging shape bias for solids can either strengthen or weaken other types of biases depending on the strength of the cues to solidity or non-solidity; further, these results make predictions about children's biased word learning over time.

Keywords: Computational models; neural networks; trajectories; word learning; shape bias.

Introduction

Computational models have proven to be an important tool for investigating many issues within cognitive development (e.g., Munakata & McClelland, 2003). Such models can provide insights about the mechanisms that underlie learning patterns seen across childhood. In the domain of word learning, various models have been used to investigate fast mapping and the taxonomic shift (Mayor & Plunkett, 2010), variability effects in learning phonetically similar words (Apfelbaum & McMurray, 2011), task effects in novel noun generalization (Samuelson, Schutte, & Horst, 2009), and word learning at different levels of abstraction (Xu & Tenenbaum, 2007). In this paper we focus on using connectionist models to examine developmental trajectories in word learning. This approach has the potential to guide novel and testable predictions about children's language development.

This paper employs a computational modeling approach to investigate the emergence of word learning biases that support early language acquisition. This approach allows us to analyze in detail how different word learning biases interact and influence one another over the course of word learning. For example, does a later emerging bias build onto and benefit from an earlier bias, or is there a period of conflict as new knowledge is assimilated with prior knowledge? Our results indicate that different biases do interact, and the emergence of one bias can either strengthen or weaken other biases depending on the strength of cues provided in the learning context . These results allow us to make predictions about the timing of children's word learning and generalization as biases emerge.

Word Learning Biases

One of the reasons that children are such skilled language learners is because of biases. In the context of word learning, biases are constraints on the range of things that children will consider in deciding what a new word refers to. Rather than assuming that any word can be used to label any item, children exhibit principled patterns of behavior in the ways in which they learn words. The main constraint we will focus on in this paper is found within the domain of noun learning: the shape bias. The shape bias refers to young children's tendency to generalize newly learned nouns to other objects based on similarity in shape (Landau, Smith, & Jones, 1988). That is, if a child is taught a novel name for a novel solid object, he or she will extend that name to other objects that match the original in shape, even if that shape match differs in texture, color, or size. Children show a reliable shape bias by 2 years of age (Samuelson & Smith, 1999).

A related phenomenon in noun learning is the material bias. While the shape bias is seen in children's generalization of labels to solid objects, the material bias concerns the labeling of non-solid substances. The material bias has been found using the same novel noun generalization (NNG) paradigm typically used in studies of the shape bias. Children taught a novel name for a novel non-solid substance tend to generalize that name to other non-solids that match the original in material rather than to non-solids matching in features like shape and size but made out of a different material (e.g., Soja, 1992; Soja, Carey, & Spelke, 1991). The material bias is typically seen slightly later than the shape bias, at 3 years of age (Yoshida & Smith, 2005).

Altogether, the evidence suggests that over the first years of life children develop preferential attention to different

features of items in noun learning, first to shape in naming solid objects and then to material in naming non-solid substances. This raises the question of whether and how these biases interact with each other. Does development of the shape bias earlier on have any impact on the material bias? Research looking at naming and generalization of other, ambiguous kinds of items hints at this possibility. For example, deformable items usually have a characteristic shape, but are also often categorized as being similar to each other in material (e.g., an item such as *towel*; Samuelson, Horst, Schutte, & Dobbertin, 2008). Previous research shows that young children categorize deformable items based on similarity in material when they are not labeled, but categorize based on similarity in shape when the items are labeled (Samuelson & Smith, 2000). Samuelson and colleagues (2008) hypothesize that this behavior represents an overgeneralization of the shape bias. This suggests that the emergence of the shape bias can influence how children learn and generalize names for other, more ambiguous kinds of items as well.

The Emergence of Biases

Although there is some debate over the origin of word learning biases (e.g., see Samuelson & Bloom, 2008), there is strong evidence to suggest the link between vocabulary growth and the emergence of word learning biases. This has been explored especially in relation to the shape bias, with several pieces of evidence suggesting interactions and feedback between attending to shape in the context of learning names for solid objects and overall word learning.

First, there is evidence that the emergence of the shape bias influences subsequent word learning. For example, Smith, Jones, Landau, Gershkoff-Stowe, and Samuelson (2002) intensively trained 17-month-old children on labels for novel shape-based categories. The children exposed to this training not only developed a shape bias earlier than is typically seen, they also showed a dramatic increase in vocabulary size compared to a control group. These results suggest that the development of the shape bias accelerates children's learning of object names outside of the lab.

Evidence for a feedback relationship between word learning and the emergence of the shape bias comes from a study by Gershkoff-Stowe and Smith (2004). In this study, children were longitudinally tested on their attention to shape in generalizing a novel label. The researchers also collected diaries tracking children's vocabulary growth. The results showed that children's attention to shape increased as the number of nouns in their vocabularies increased.

Together these studies suggest an interesting pattern of interactions between language acquisition and the emergence of word learning biases. As children add more nouns to their growing vocabularies, they show an increasing preference to attend to shape in the context of naming solid objects. This preference to attend to shape in turn facilitates subsequent word learning. This leaves open the question of how different biases may interact over the course of language development. Modeling word learning offers a unique way to investigate this question.

Modeling Word Learning & Biases

Computational models of word learning have been used to investigate the conditions that support word learning biases. For example, Colunga and Smith (2005) trained a network with a vocabulary structure of half solid objects characterized by shape and half nonsolid objects characterized by material—a vocabulary structure which should directly promote the development of shape and material biases. Results of the virtual analog of the NNG task confirmed this prediction in that the network showed a shape bias for solids and a material bias for nonsolids. In a second simulation, they trained the same network on a realistic early vocabulary, structured like that of a typical 30-month-old. The earliest nouns that children typically learn are dominantly comprised of solid objects characterized by shape (e.g., *ball*, *spoon*), and include fewer non-solid substances characterized by material. The typical early noun vocabulary also includes types of items that can be characterized by both shape and material. When trained on this more complex vocabulary structure, the network again treated novel test items in ways consistent with shape and material biases. Colunga and Smith (2005) also reported behavioral data with young children confirming the predictions of this network. More recently, Colunga and Sims (2011) used the same kind of network to successfully predict differences in novel noun generalization patterns between early- and late-talker children.

These studies show that computational modeling is a powerful tool for exploring the emergence of word learning biases. However, no one has yet used this approach to investigate the relationships between different biases as they emerge over the course of word learning. Our approach offers a new perspective by modeling different word learning biases together on a developmental timescale.

Approach and Overview

Our approach is to train a network on a typical early child vocabulary and then test it on generalization of three different kinds of items over the course of word learning. The goal is to see how the shape bias emerges over word learning and how the emergence of the shape bias impacts generalization performance on other kinds of items. To test for a shape bias, we implement a virtual NNG task by exposing networks to novel solid items and seeing if they treat those alike in shape more similarly than those alike in material (as in Colunga & Smith, 2005). We also test the network on two types of novel non-solid items: simply-shaped, clearly material-based non-solids and complex-shaped, ambiguous non-solids. An example of a simply-shaped nonsolid would be a glob of paint. While it can take on slightly different shapes, it always maintains its "blob-like" shape. Simply-shaped non-solids are inherently material-based because there is more variation in the material than in the shape, making the material of an item a

more useful cue in classification. Complex-shaped, ambiguous non-solids, on the other hand, can take on virtually any shape, e.g. paint smeared in the shape of a peace symbol. In this case, shape and material have equal variability and either one could be a cue to classifying an item. We explore whether the network develops a material bias for these types of non-solids, and whether that development depends on the emergence of the shape bias for solids.

Simulation

Method

Computational models use the Leabra algorithm (Local, Error-driven and Associative, Biologically Realistic Algorithm), which combines Hebbian and error-driven learning. Weights are adjusted based on correlations between activation units and feedback of errors through back propagation (O'Reilly et al., 2012).

Architecture The architecture is adapted from Colunga and Smith (2005) and is implemented as shown in Figure 1. Words are represented discretely and are input on the Word Layer. Referents are represented as distributed patterns over several dimensions on the Perceptual 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 Perceptual Layer, with 12 units for shape and 12 units for material. Solidity is represented discretely; one unit stands for Solid and another for Non-Solid. Finally, there is a 25 unit Hidden Layer that is connected to all the other layers and to itself. The Hidden Layer serves as the bridge between the Word Layer and the Perceptual Layer and it is where learning occurs. Learning progresses as internal representations, or weights, update and form links between the other two layers.
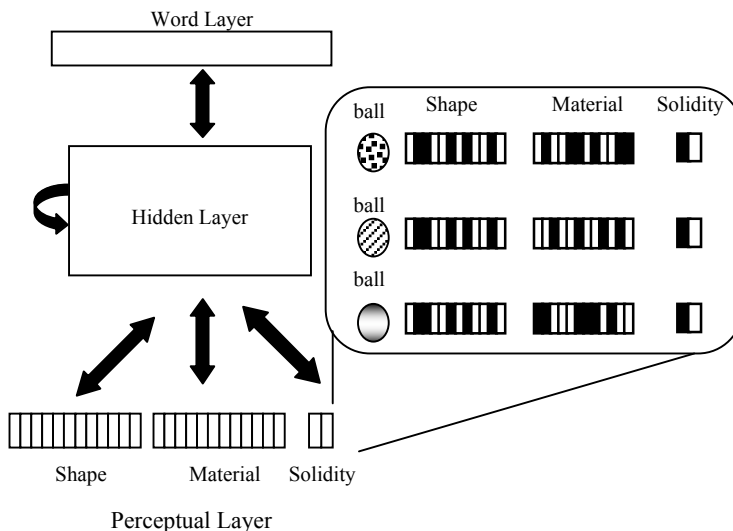

Figure 1: Architecture of the network and example input patterns.

Table 1: Noun category proportions used to create the input vocabulary structure. Beneath each proportion is an example noun belonging to that category.

	Shape	Material	Both
Solid	52% *ball*	10% *cheese*	12% *crayon*
Non-solid	4% *bubble*	16% *glue*	6% *jeans*

Input Patterns The input patterns consisted of training and testing patterns. The training patterns were structured to capture the same correlational structure as the vocabulary of a typical 30-month-old child (Fenson et al., 1993). The structure of this training input was based on that used in Colunga and Smith (2005). Using adult judgments, nouns were categorized by both solidity (either solid or non-solid) and characteristic feature (either shape, material, or both). The structure of the typical early noun vocabulary could then be expressed as proportions of each type of category. See Table 1 for the 6 categories and proportions used in the current study. The network in this study was trained to learn 50 noun representations, designed to have approximately the category structure of the typical 30-month-old's vocabulary. Note that these 50 noun representations are word inputs to the network and do not necessarily correspond to 50 nouns that a child learns.

The test input patterns consisted of three kinds of items represented along the Perceptual Layer: solids, simply-shaped non-solids, and complex-shaped non-solids. Test patterns were made up of triplets of novel items: an exemplar and two choice items.

For solid test items, this pattern was instantiated by manipulating activation across all shape and material units of the network. For clear non-solids, half of the shape units were kept constant to capture the fact that non-solid substances are typically seen in simple shapes like smears and splashes.[1] That is, these simply-shaped, clear non-solids had some variability in the shapes they could take on, but also had some imposed limitations. Finally, the ambiguous non-solid test patterns were represented by manipulating activation across all of the shape and material units. This type of non-solid test pattern is called ambiguous because, unlike other non-solids the network was exposed to, they can take on more complex shapes and thus be construed as having a characteristic material or a characteristic shape. Therefore, testing involved generalization to novel solid and simply-shaped, non-solid items as well as generalization to a new, ambiguous type of non-solid item.

Progression of Word Learning To chart the course of bias development, we tested the network at multiple points throughout word learning. Weights were recorded at initialization, every five words from 0 to 30 words learned, and every ten words from 30 to 50 words learned. The

---

[1] This method for simplifying the shapes of non-solids was also used in the training input patterns.

endpoint of learning was recorded as either asymptotic performance of learning all 50 words, or at the end of 500 epochs of training. This resulted in ten checkpoints along the trajectory of word learning. Although other measures such as duration of training could capture how much a network has learned, we used number of words learned because this is the key factor driving the development of biases.

Training. On each trial of training, a word was paired with a referent. The patterns associated with each word were determined based on which noun category that word was meant to represent. For example, a word for a solid item characterized by shape (like a ball; see Figure 1) should be used to label things that are like each other in shape but differ from each other in material. To simulate this pattern, we randomly selected an input vector to represent, for example, ball shape. On individual training trials, we paired that shape pattern with the label *ball* and a randomly selected material pattern. Therefore over multiple training trials, a word for a solid item characterized by shape would be represented by the same shape but different material patterns. We did this for each of the 50 nouns in the training set. This part of the simulation was intended to put into the network the lexical knowledge that a typical child would bring to the laboratory NNG task.

Testing Following training, the next step was to identify what sort of word learning biases each run of the network had developed. We addressed this question with a virtual version of the NNG task. We presented the network with three NNG tasks, one for each of the three types of items: solids, clear non-solids, and ambiguous non-solids. On each test trial of these tasks, we presented the network with a triad of novel entities (one at a time) on the perceptual layer. The triad consisted of an exemplar and two choice items, one matching the exemplar in shape only and one matching in material only. The only difference between the trials for the clear non-solids and those for ambiguous non-solids was that for clear non-solids, the material-matching choice item had a simplified pattern along the shape layer, as discussed in the input patterns section. 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 this same relationship is *less*, then the internal representations highlight the material of the items. We used these similarities along with Luce's choice rule (Luce, 1963) to calculate probability of choice in order to predict performance in the NNG task.

Results

We averaged over 10 runs with different initial random weights. First we looked at the network's test output across the entire course of learning. As shown in Figure 2, the network preferred the shape match choices at test to different extents depending on both the solidity of the item presented and the size of its vocabulary at that point. For example, before training, the network showed no preference for either the shape match or the material match test choice for solid items, tending to choose the shape match about half the time ($M = .50$, $SD = .02$). In contrast, by the end of training the network chose the shape match for solid items about three quarters of the time at test ($M = .78$, $SD = .04$).

On the other hand, while the network started at a similar state for complex-shaped, ambiguous non-solids, it showed a different preference by the end of training. Specifically, the network began with no preference for shape or material, but developed a shape preference for these non-solids, albeit to a lesser extent than it did with solid items ($M = .62$, $SD = .03$). This pattern shows an overgeneralization of the shape bias to this particular kind of non-solid item.
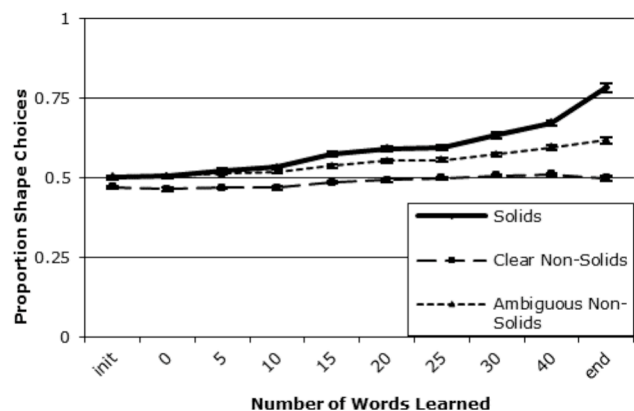

Figure 2: Mean proportion of shape match choices at test for each kind of test item. Error bars represent standard error.

Still another pattern was seen in how the network treated simple-shaped, clear non-solids at test. In this case, the network actually began with a preference for material, choosing the shape match test choices slightly but significantly less than 50% of the time ($M = .47$, $SD = .01$, $t(9) = -8.31$, $p < .001$). This inherent material bias was present at initialization due to the input structure of the clear non-solids items. Recall that for this type of item, half of the shape units were kept constant across all of the input. Because of this, at the time of initialization the networks had less information about variations in shape on which to base representations. Instead, the networks harnessed the relatively richer material information immediately available about clear non-solid items, and thus showed a slight preference for this feature initially. However, by the end of learning the initial preference for material was gone ($M = .50$, $SD = .02$). What caused the network to lose this early preference for material in the context of clear non-solids? And more specifically, did it have anything to do with what

was concurrently being learned about solid items? To get at this question, we next focused our analysis on the time window in which the shape bias emerged in learning.
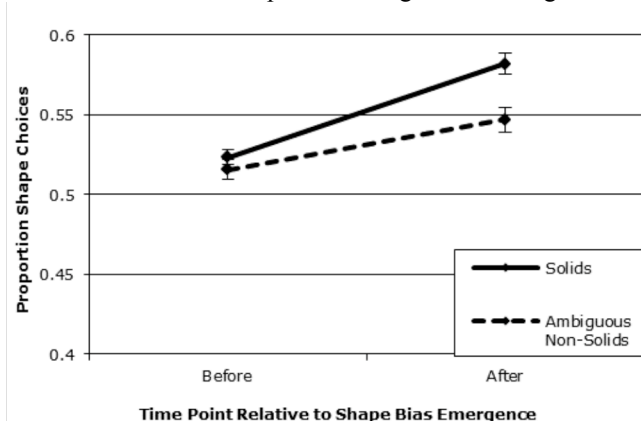

Figure 3: Mean proportion of shape match choices at test for solids and ambiguous non-solids immediately before and after the emergence of the shape bias for solids.

Identifying the point in learning at which the network first developed a shape bias involved several steps. First we examined the output of each of the ten runs and picked out when in learning the network chose the shape match test choice for solid items greater than 55% of the time.[2] This happened on average around the time the network had learned 15 words from the training set. The next step was to identify how this point of emergence of the shape bias in solids affects the network's behavior concerning the two kinds of non-solids. We isolated the closest time points in learning both preceding and following the emergence of the shape bias. Using this focused time window, we examined how the network treated each type of non-solid item in relation to solid items as the shape bias for solids emerged.

To examine the interaction between the emerging shape bias and the network's choices among complex-shaped, ambiguous non-solid test items, proportion of shape choices were submitted to a 2 (test item type: solid or ambiguous non-solid) × 2 (time point in learning: before or after the emergence of the shape bias) mixed design analysis of variance (ANOVA).[3] This analysis yielded a significant main effect of time point, $F(1, 30) = 13.41, p = .001$, with no other significant effects. This shows that as the shape bias for solids emerges in the course of learning, a similar bias develops for ambiguous non-solids (see Figure 3), suggesting an overgeneralization of the shape bias to ambiguous non-solids. Further, this shape bias for both types of items increases over the time window of interest.

Next we examined the interaction between the emerging shape bias and the network's choices among simply-shaped, clear non-solid test items. Proportion of shape choices were

submitted to a similar 2 (test item type: solid or clear non-solid) × 2 (time point) ANOVA. This analysis yielded main effects of both test item type ($F(1, 30) = 25.47, p < .001$) and time point ($F(1, 30) = 11.14, p < .01$). These main effects are qualified by a significant interaction between test item type and time point in learning, $F(1, 30) = 8.73, p < .01$. As can be seen in Figure 4, as the shape bias for solids emerges, the network's preference for shape choices for clear non-solids also increases, however the nature of these changes differs between item types. In fact, the network shows a preference for material test choices for clear non-solids, even just after the emergence of the shape bias for solids. Although this preference for material diminishes somewhat over the time window in question (as shown by an increase in proportion of shape choices), it does not do so at a rate proportional to the growth of the shape bias for solids. This suggests that the emergence of the shape bias may have a slight diminishing influence on the material bias for clear non-solids.
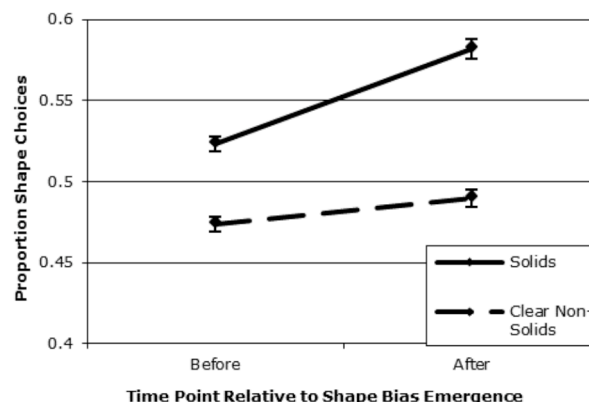

Figure 4: Mean proportion of shape match choices at test for solids and clear non-solids immediately before and after the emergence of the shape bias for solids.

Finally, to check that these patterns were specific to the time window surrounding the emergence of the shape bias, we ran the same analyses for the immediately following time window. The only effect that reached significance was a main effect of item type between solid and clear non-solid test items, $F(1, 32) = 37.62, p < .001$. As in the above analysis, this reflects an ongoing difference in the extent of shape choices that the network made in the context of these two types of items. The fact that no other effects were significant supports the argument that the preceding results are specific to the time window in learning immediately surrounding the emergence of the shape bias.

Discussion

The current study investigated the dynamics involved in the development of the shape bias for solid items and learning about other kinds of items in early language acquisition. We found that the emergence of the shape bias for solids led to either an overgeneralization of the shape bias or a diminishment of the material bias, depending on the

---

[2] 55% was chosen as the threshold for emergence because at this value the network chose shape matches reliably greater than chance across the ten runs, $t(9) = 12.54, p < .001$.

[3] This type of analysis was chosen following Colunga & Smith (2005) in their analyses of similar networks.

strength of the cues to solidity or non-solidity provided by the test items. depending on the type of test item used.

First, for complex-shaped, ambiguous non-solid test items, we observed an overgeneralization of the shape bias. This is consistent with findings that 3-year-old children overgeneralize the shape bias in extending names to deformable items (Samuelson et al., 2008). This suggests that for ambiguous categories, the word learning bias that is already established, typically the shape bias, takes precedence and guides generalization.

Second, for simply-shaped, clear non-solid test items, the material bias diminished as the shape bias for solids developed, although it did so at a slower rate compared to the growth of the shape bias. This suggests that the material bias for clear non-solids was more resistant to the influence of the developing shape bias, perhaps in part due to the inherent material preference for these items. This inherent bias was due to the structure of the input itself, but a similar early material bias has also been seen in children if they are tested with simply-shaped non-solid substances on the NNG task (Colunga & Smith, 2005).

The mechanisms driving the observed changes in bias learning and development in our networks speak directly to possible mechanisms in children. The current work suggests that the network can mainly develop one bias at a time, which would explain why as the shape bias for solids takes off, the network also shows a growing preference for shape for the two other types of items. If this is the case, we would predict that as the network continued to train on more words, the material bias for non-solids would come online (as seen in children) and cause a dip in the established shape bias for solids. However, by this time the shape bias for solids would be well established and thus largely resilient against the network's shift to focusing on material. After enough exposure to the vocabulary structure, we would expect the network to have acquired both a shape bias for solids and a material bias for non-solids, as is seen in children later in language development.

This work could be extended to make predictions about and map to specific places of bias emergence in individual children's word learning. By using longitudinal MCDI data from individual children, one could use networks to analyze the development of language biases in early- and late-talkers (as done in Colunga and Sims 2011), looking specifically at how the emergence of certain word learning biases affects other biases over time. These biases could show different patterns of interaction among early- and late-talkers. In sum, this work opens the door for further modeling and can make novel, testable predictions about the development of children's word learning.

Acknowledgments

References

Apfelbaum, K.S., & McMurray, B. (2011). Using variability to guide dimensional weighting: Associative mechanisms in early word learning. *Cognitive Science, 35,* 1105-1138.

Colunga, E., & Sims, C.E. (2011). Early talkers and late talkers know nouns that license different word learning biases. In L. Carlson, C. Hoelscher, & T. Shipley (Eds.), *Proc. of the 33rd Annual Conference of the Cognitive Science Society* (pp. 2550-2555). Austin, TX.

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

Fenson, L., Dale, P., Reznick, J.S., Thal, D., Bates, E., Hartung, J., et al. (1993). *The Macarthur Communicative Developmental Inventory: User's guide and technical manual.* San Diego, CA: Singular Publishing.

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

Luce, R.D. (1963). Detection and recognition. In R.D. Luce, R.R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (pp. 1-41). New York: Wiley.

Mayor, J., & Plunkett, K. (2010). A neurocomputational account of taxonomic responding and fast mapping in early word learning. *Psychological Review, 117,* 1-31.

Munakata, Y., & McClelland, J.L. (2003). Connectionist models of development. *Dev. Science, 6,* 413-429.

O'Reilly, R. C., Munakata, Y., Frank, M. J., Hazy, T. E., (2012). *Computational Cognitive Neuroscience*. Wiki Book, 1st Edition. URL: http://ccnbook.colorado.edu

Samuelson, L.K. & Bloom, P. (2008). Special section: What counts as an explanation of development: The shape bias as a case study. *Developmental Science, 11,* 183-184.

Samuelson, L.K., Horst, J.S., Schutte, A.R., & Dobbertin, B.N. (2008). Rigid thinking about deformables: Do children sometimes overgeneralize the shape bias? *J. of Child Language, 35,* 559-589.

Samuelson, L.K., Schutte, A.R., & Horst, J.S. (2009). The dynamic nature of knowledge: Insights from a dynamic field model of children's novel noun generalization. *Cognition, 110,* 322-345.

Samuelson, L.K., & Smith, L.B. (2000). Children's attention to rigid and deformable shape in naming and non-naming tasks. *Child Development, 71,* 1555-1570.

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

Soja, N.N. (1992). Inferences about the meaning of nouns: The relationship between perception and syntax. *Cognitive Development, 7,* 29-45.

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

Xu, F., & Tenenbaum, J.B. (2007). Word learning as Bayesian inference. *Psych. Review, 114,* 245-272.

Yoshida, H. & Smith, L.B. (2005). Linguistic cues enhance the learning of perceptual cues. *Psych. Science, 16,* 90-95.