Learning individual words and learning about words simultaneously

Sylvia Yuan (shyuan@berkeley.edu)
Department of Psychology, University of California, Berkeley
Berkeley, CA 94720 USA

Amy Perfors (amy.perfors@adelaide.edu.au)
School of Psychology, The University of Adelaide
Adelaide, SA 5005 Australia

Josh Tenenbaum (jbt@mit.edu)
Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology
Cambridge, MA 02139 USA

Fei Xu (fei_xu@berkeley.edu)
Department of Psychology, University of California, Berkeley
Berkeley, CA 94720 USA

Abstract

Children are guided by constraints and biases in word learning. In the case of the shape bias—the tendency to extend count nouns by similarity in shape—previous findings have revealed that learning plays an important role in its development (e.g., Smith et al., 2002). Some have proposed that children acquire inductive constraints like the shape bias by making inferences about observed data on multiple levels of abstraction (e.g., Smith et al. 2002; Kemp et al., 2007). The current study provides support for this hypothesis by demonstrating that preschoolers can rapidly and flexibly form overhypotheses about the role of arbitrary features, not just shape, in determining how words refer to object categories. This work suggests that when learning individual words, children are also learning *about* words simultaneously. Children's ability to "learn to learn" may have implications for the origins of learning biases in many different cognitive domains, not just in language learning.

Keywords: word learning; shape bias; inductive constraints; overhypothesis; rational inference

Introduction

Much of what we know about the world depends on inductive learning—inferring an underlying general principle based on limited data. Induction is not a trivial problem: in principle, there is an infinite set of possible generalizations that can be made from the same observed examples (Quine, 1960). For example, in the domain of word learning, from hearing the word 'dog' while observing a situation involving the presence of a dog, a learner could hypothesize that the word refers to *fur, cute, tail, dog that is alive*, or *undetached dog parts*, among a potentially infinite set of possible meanings.

Learning must therefore be constrained by biases of some sort (Keil, 1981; Markman, 1989). Children rely on inductive constraints in many cognitive domains, such as word learning (e.g., Landau, Smith, & Jones, 1998; Markman, 1989) and physical and psychological reasoning (Baillargeon, 2008; Carey & Spelke, 1996; Gergely & Csibra, 2003). Given the early appearance of these constraints, it seems conceivable that some may be innately given. It is possible, however, that some of the early cognitive biases might be learned. For example, 1.5-year-olds can be trained to exhibit a shape bias in word learning, extending a newly-learned label to a similarly-shaped object (e.g., Smith et al., 2002; see also Samuelson, 2002).

The acquisition of inductive biases continues through childhood and adulthood and takes place not only in the domain of word learning. Based on only a few examples, toddlers and preschoolers can learn the dimension used in classification and apply this knowledge to sort new objects into new categories (Macario, Shipley, & Billman, 1990; Mareschal & Tan, 2007). From observing causal relations of an initial set of objects, preschoolers and adults form abstract causal schemata and use them to make inferences about the behaviors of newly-encountered objects (Kemp, Goodman, & Tenenbaum, 2010; Lucas, Gopnik, & Griffiths, 2010). In all of these cases, learners rapidly acquire abstract knowledge of some form that helps them readily learn about novel items or situations based on sparse data. How do learners acquire knowledge that guides subsequent learning? How do they "learn to learn"?

A constraint on learning, whether in the form of perceptual biases (e.g., shape bias) or principles or systems of relations (e.g., causal schemata), is a type of abstract knowledge specifying how things work in general, going beyond individual instances. Having such a constraint thus requires learners to represent knowledge on multiple levels; the constraint itself is a form of higher-level knowledge, or overhypothesis (Goodman, 1955). In the case of the shape bias, the learner may have some lower-level knowledge that objects labeled as 'ball' are all spherical, but do not seem to all be white or plastic; this first-order knowledge is about each individual category. Having the shape bias means that the learner also has some higher-level knowledge that objects that share the same name share the same shape, but

not color or material; this second-order knowledge is an overhypothesis about how categories in general are organized. This higher-level knowledge allows learners to be able to readily extend a newly learned count noun to new instances.

Several computational models have been proposed to explain how the acquisition of higher-level knowledge can account for the emergence of some inductive constraints such as shape bias. Some models provide a more mechanistic account of learning (e.g., Colunga & Smith, 2005), while others focus on the computational principles underlying the acquisition of multi-level knowledge (e.g., Kemp, Perfors, & Tenenbaum, 2007; Perfors & Tenenbaum, 2009).

The primary focus of this paper is to examine what factors influence children's acquisition of higher-order abstractions, and to what extent they are able to use such abstractions to guide subsequent learning. Previous work indicates that toddlers and preschoolers can rapidly achieve higher-order generalizations about object dimensions (e.g., shape) in word extension and in object categorization, based on very little input (e.g., Smith et al. 2002; Macario, et al., 1990). Here, we aim to explore the flexibility and limits of such rapid higher-order learning in children—in particular, whether children are capable of learning higher-level knowledge about arbitrary features. Our goal is to shed light on the kinds of mechanisms children engage in for acquiring higher-level abstractions and the experimental paradigm we develop may be helpful in the future to discriminate among different models of higher-level learning.

This study explores whether children's rapid acquisition of overhypotheses requires the target feature to be already salient (like shape or color). If inductive constraints such as the shape bias emerge as a general mechanism of acquiring higher-level knowledge, then we would expect children to be able to form higher-level abstractions over dimensions that are not shape. That is, given appropriate statistical regularities in the input, children should be able to acquire overhypotheses over dimensions that are less salient than shape and perhaps even arbitrary dimensions they have never been exposed to. Recent findings suggest that adults can do so, forming overhypotheses rapidly over arbitrary dimensions in learning artificial object categories (Perfors & Tenenbaum, 2009). It is unclear whether children can do so quickly with arbitrary and completely novel features.

To test this, we asked two questions. First, can children form higher-level generalizations about the relevance of arbitrary dimensions in a word-extension task? Second, can they do so rapidly, on the basis of small amounts of data – as quickly as they form first-order generalizations? The task was modeled after a similar task with adults (Perfors and Tenenbaum, 2009). Children were shown categories of animals with novel labels for each category, where the animals were organized according to symbols on some body parts. Afterwards, they were asked to extend a trained name (first-order generalization) and a novel name (second-order generalization).

A secondary goal of the current study is to examine the kinds of input that affect children's ability to form higher-level generalizations in word learning. While Smith et al. (2002) provided evidence that toddlers can achieve higher-level generalizations about the role of shape in word-to-object mapping, it is unclear what aspects of the input most influenced this. Did the rapid higher-order learning occur because shape was such a highly coherent feature in the object categories children were exposed to? In the real world, there is certainly noise in the input—e.g., chairs are not always shaped in the same way. As a result, we would expect children to be able to achieve higher-level abstractions when the relevant feature is not 100% coherent. Can they indeed do this? We are also interested in exploring how category structure—in terms of the number of categories and the number of total items—might influence higher-level learning. We addressed these questions in this study by varying children's input in terms of feature coherence and category structure.

Method

Participants

Sixty 4- to 6-year-olds (Mean 58.5 months; range 48.3-79.8; 37 boys, 23 girls), all native speakers of English, participated. Twelve additional children were eliminated due to refusal to participate (2), inattentiveness (3), or side bias (7, see Procedure below). Families were recruited from the Berkeley area and surrounding communities.

Materials

The stimulus materials consisted of images of artificially constructed animal categories, presented on a computer laptop. Within each trial, all instances of the animals in the training and test phases were of the same overall geometric shape (see Figure 1). For each instance of animal, each of four particular body parts (e.g., hump, tail, front leg, and hind leg) contained a different symbol. Two of these body parts each contained a symbol that was shared among animals of the same training category (e.g., 'δ' on the hump and 'V' on the hind leg, in the trial displayed in Figure 1); these were the *relevant* features for categorization. The other two body parts (e.g., tail and front leg) each contained a symbol that varied among animals of the same training category; they were the *irrelevant* features.

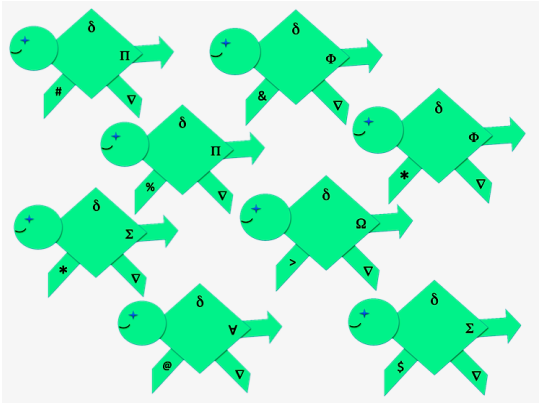
The animals from different trials (see Design below) differed considerably in their appearances in terms of color scheme and morphology. Moreover, the relative locations of the relevant and irrelevant features also varied across trials. For example, in a different trial, the animals were gray and looked bear-like standing upright, with the front left paw and right ear as the relevant features, and the tummy and the right hind paw as the irrelevant features. This was done to increase stimulus diversity and thus to minimize perseveration by children across trials. A different set of novel names was used to label the training categories of animals in each trial.

Design

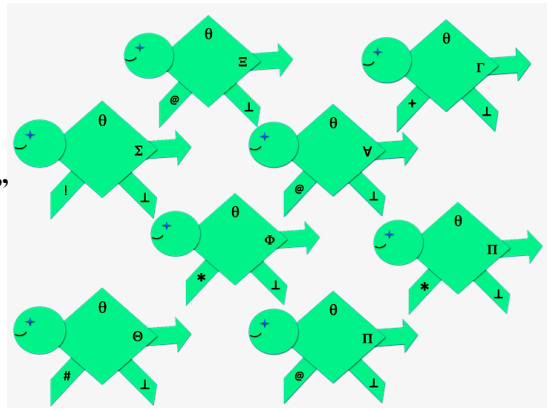
All children received 2 orientation trials and 6 critical trials.

a. Training phase

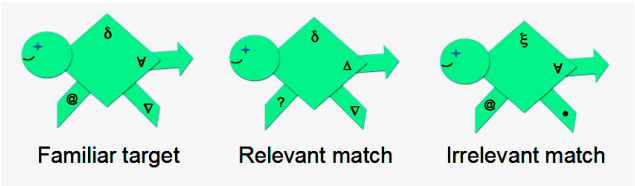
"bilarks"



"sarlins"



b. First-order generalization



c. Second-order generalization

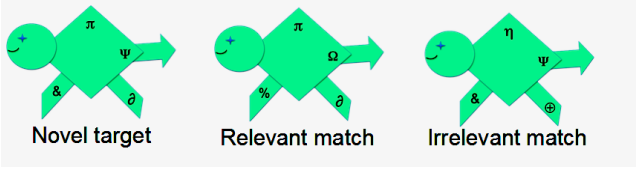


Figure 1: Example stimuli for a critical trial

Trial structure Each critical trial consisted of 3 phases: training, the first-order generalization test, and the second-order generalization test. In the training phase, children were shown a number of categories of animals and told their names (Figure 1a). After the training phase, each child was tested on first-order and second-order generalization questions. In the first-order generalization test, children were shown an animal instance they had already seen in the training phase (familiar-target), and were asked which of two other novel animal instances was of the same name (Figure 1b). One choice item shared the same marks on the relevant features (relevant match), and the other choice item shared the same marks on the irrelevant features (irrelevant match). The second-order generalization test was structured similarly, except that the target was a novel animal instance, with feature values children had not seen before (novel-target; Figure 1c). Half of the children received the first-order generalization test before the second-order generalization test for all trials, whereas the other half received the tests in the reverse order.

Trial type The trials varied in the coherence of category features (75% or 100%) and in category structure (16-4, 16-2, and 8-2). In the 100%-coherence trials, all animals of each training category shared the same feature value on each of the relevant features; in the 75%-coherence trials, only three quarters of the animals did. The 3 types of category-structure trials varied in the number of training total items in training and the number of training categories: the 16-4 trials presented 16 total items in 4 categories; the 16-2 trials with 16 total items in 2 categories; the 8-2 trials with 8 total items in 2 categories. Each child received a total of 6 critical trials, from crossing category features (2) and category structure (3). Figure 1 provides an example of a 100%-coherence and 16-2-category-structure trial.

Procedure

The children were told that they would be playing a sticker game involving the computer. Each child was seated in front of a laptop computer, with the experimenter sitting next to the child and using a mouse to advance the slides on the computer. The experimenter explained to the child that the child would see some animals that had marks on their body parts, be told what the different animals were called, be asked to point to some animal every once in a while, and receive a sticker each time they responded.

The experiment then proceeded with 2 orientation trials and 6 critical trials. The orientation trials were designed to familiarize the children with the task. They were similar to the critical trials in structure, consisting of a training phase and a single test.

In the training phase of the first orientation trial, the child saw 2 novel categories of animals, on separate slides. The animals looked identical, except that animals of the first category had the symbol '#' on their tummy while animals of the second category had the symbol '@' on their tummy. While each category of animals was displayed, children were told a novel name for the animals (e.g., "Look! These are *tomints*") and then saw red circles appear around the marks ("And they have a mark right there"). Each category of animals was accompanied by a different novel name (*tomints*, *lampiles*).

In the test of the first orientation trial, the child saw shrunk images of the training categories of animals on the top of a slide and were reminded of what they were (e.g., "So, we have the tomints and the lampiles"). The child then was presented with the target item in test—one example animal of the first category—accompanied by its name ("Remember this one? This is a tomint"). Next, the child saw two animals side by side, one identical to the target (match) and one identical to the target but without the mark on the tummy (mismatch). The child was asked to choose one that had the same name as the target animal (e.g., "Can you point to another tomint?"). The child was given positive feedback for her correct choice and received a sticker. In the rare case that a child chose the mismatch in an orientation trial, the experimenter would repeat the trial until the child correctly chose the match. The second orientation trial proceeded in the same manner, but with a different set of animals, different marks on a different body part, and different novel names.

Each of the critical trials proceeded in the same manner as with orientation trials, except that each critical trial included two tests (first- and second-order generalization), and that children received positive feedback regardless of which choice item they chose on tests. In the first-order generalization test, children saw the familiar-target and were reminded of its name (e.g., "Remember this one? This is a bilark"), then were asked to choose between the relevant match and the irrelevant match as another instance of the same category ("Can you point to another bilark?"). In the second-order generalization test, the children were introduced to the new target paired with a new novel name (e.g., "Look at this new one. This is a morple"), then were asked to choose between the relevant match and irrelevant match as another example of the same name ("Can you point to another morple?").

The 6 critical trials were presented in pseudo-randomized order, with no more than 2 trials of the same coherence level in a row. Across the trials, children were presented with a total of 12 generalization tests. The side of the relevant match was never the same for more than 3 times in a row. Children who pointed to the same side 11 or more times out of the 12 tests were eliminated as exhibiting side bias.

Coding

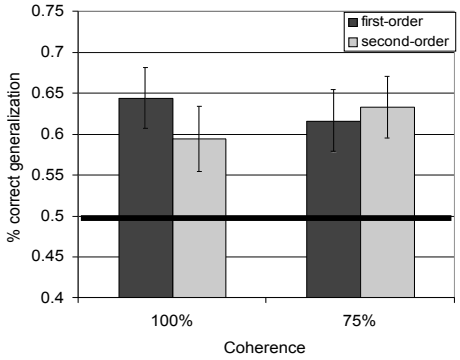
The percent generalization was calculated as the number of times the child chose a relevant match out of the total times the child chose either a relevant or an irrelevant match.

Preliminary analyses of overall percent generalization revealed no effects of sex, age group (younger versus older half), or whether children received the first-order or second-order generalization first. The data were therefore collapsed across sex, age group, and test order.

Results

As predicted, children's first-order and second-order generalization did not differ significantly ($t(59) < 1$). In the first-order generalization tests, children generalized the trained name to new instances by the relevant features 63% of the time (SD = 19), reliably different from chance (chance = .5; $t(59) = 5.34$, $p < .001$). Similarly, in the second-order generalization tests, children generalized the novel name to new instances by the relevant features 61% of the time (SD = 18), also reliably different from chance ($t(59) = 4.90$, $p < .001$). Thus, children were able to generalize a new novel name to new instances as soon as they were able to generalize a trained name.

a. Generalization according category coherence



b. Generalization according to category structure

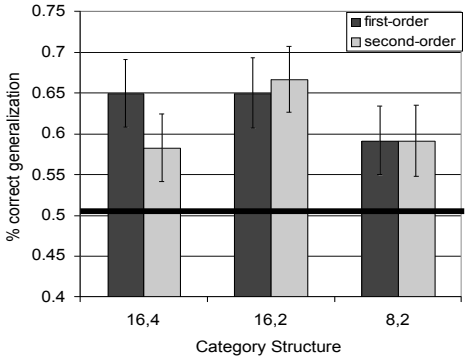


Figure 2: Children's generalization performance by coherence (2a) and category structure (2b). Error bars represent standard errors. Generalization was above chance (50%), but did not differ significantly by coherence, category structure, or level (first vs second-order).

Children's generalization did not appear to differ with coherence level or with category structure. A 2 (coherence: 75%, 100%) by 2 (order-of-generalization: first-order, second-order) repeated-measures ANOVA revealed no effects of coherence, order of generalization, or interaction of the two factors (F's < 1.03). Similarly, a 3 (category structure: 16-4, 16-2, 8-2) by 2 (order-of-generalization: first-order, second-order) repeated-measures ANOVA revealed no effects of category structure, order of generalization, or interaction of the two (F's < 1). As Figure 2 indicates, children performed reliably above chance on first-order and second-order generalization tests for both 75%- and 100%-coherence trials (Figure 2a; t's > 2.37, p's

< .03), and for all three category-structure trials (Figure 2b; t's > 2.00, p's < .05).[1]

Discussion

These findings show that 4- to 6-year-olds were able to rapidly acquire an overhypothesis about the role of arbitrary features (not just shape) governing how words are used to refer to object categories. With as few as two categories and eight items in the training input, children in word-extension tests were able to make abstract (second-order) generalization as soon as they were able to make first-order generalizations. That is, in a matter of one to two minutes of experience with a small set of novel object categories and novel names, children could quickly learn an abstract commonality across the categories and the names, and immediately apply this abstract knowledge to guide subsequent learning. Moreover, they could do so even when there was noise in the input in terms of how coherently the training categories were organized by the relevant feature.

Although this study was somewhat limited in its power due to the fact that our children were presented with relatively few trials each, our results do suggest that children can flexibly and rapidly acquire overhypotheses. These findings imply that as children learn individual words, they are simultaneously learning abstract knowledge about words in general as well.

Yet how is such higher-order knowledge acquired? There are different proposals with regards to overhypothesis learning. In what is known as the Attentional Learning Account, Colunga and Smith (2005) have proposed that children acquire the shape bias by detecting regularities in the input. In particular, children first detect associations among solid objects, count noun syntax, and objects categories organized by shape. These associations form the basis for learning about the relations between specific words and specific categories (first-order knowledge), and are eventually followed by the emergence of abstract knowledge about relations between words and categories in the abstract (second-order knowledge).

Recently, others have proposed that overhypotheses can be learned via a rational inferential mechanism, as captured by a hierarchical Bayesian model (Kemp et al., 2007; Perfors, Tenenbaum, Griffiths, & Xu, in press; Xu, Dewar, & Perfors, 2009). The idea is that as a learner receives data, she makes inferences and updates hypotheses on multiple levels of abstraction *simultaneously*. For example, given each additional example of a ball labeled as 'ball,' the learner may, at the first-order level, value more the hypothesis that objects named 'ball' are round and value less the hypothesis that they are white. At the same time, each instance of a 'ball' also contributes to inferences on the second-order level, allowing the learner to give increasing weight to the hypothesis that objects given the same name share the same shape, and less weight to the hypothesis that objects of the same name share the same color.

These two proposals — the Attentional Learning Account and the Rational Hierarchical-inferential Approach — are similar in some respects. First, both proposals consider the role of statistical regularities in the input in shaping the kind of bias that emerges; the shape bias emerges not because it is a privileged perceptual dimension in the first place, but because of the correlations between shape-based categories and word usage. Second, both proposals construe the acquisition of bias as children arriving at some higher-order abstraction. The proposals differ with respect to the kinds of mechanisms and principles that allow the learner to make the leap from statistical regularities in the input to higher-order knowledge. While the Attentional Learning Account focuses on bottom-up associative processes, the Rational Hierarchical-inferential Approach focuses on the principles of multi-level inferences and evaluation of hypotheses. Given that each instance of data contributes to both lower and higher-order knowledge, the Rational Hierarchical-Inferential approach thus predicts that the learner can rapidly arrive at second-order knowledge as soon as (or even before) they obtain first-order knowledge (Kemp et al., 2007), without first needing to learn about many instances and categories on the lower level.

Our findings are consistent with the Rational Hierarchical-Inferential approach, which predicts that abstract inductive biases may be acquired quite rapidly, on the basis of relatively sparse data. The simultaneous inference-making on multiple levels allows the learner to quickly acquire abstract knowledge that goes beyond that given input and, importantly, guides subsequent learning. It is theoretically possible that first- and second-order knowledge are not acquired simultaneously, but *are* acquired so rapidly successively (i.e., after only 8 to 16 objects) that they appeared simultaneous in our data. Still, this is unlikely: performance, though better than chance, was not close to ceiling, implying that if learning is not simultaneous, it is only distinguishably different at the very earliest stages.

Aspects of the current results are also consistent with the Attentional Learning Account. While the rapidity of learning and the simultaneous first- and second-order generalizations might be unexpected, the ability to learn an overhypothesis involving arbitrary features, given the statistical regularities in the input, is indeed predicted by the attentional learning approach. One possible way to examine the differences in the learning mechanisms is to examine how learners might be sensitive to not just the input data, but to how the data is generated (e.g., Xu & Tenenbaum, 2007)—one key feature of rational learners.

While the children demonstrated their ability to acquire overhypotheses despite noise in the input (as evident in their performance on the 75%-coherence trials), it was somewhat surprising that they did not perform reliably worse on the

[1] When the 7 children with the side bias were included, all analyses gave the same result except for one that was no longer significant (comparing 1st-order generalization on 8-2 trials against chance).

75%-coherence trials, compared to the 100%-coherence trials. This lack of a coherence effect contrasts with a previous finding with adults, who are influenced by feature coherence in their overhypothesis learning in a category-learning task (Perfors & Tenenbaum, 2009). It is possible that there was not enough power in the current dataset, given that the children received many fewer trials compared to the adults in Perfors and Tenenbaum's study (6 for children, 30 for adults). It is also possible that coherence would affect learning more when the learner does not have precise information about category membership of the training instances, as with adults. Future work will explore this possibility with children.

The lack of an effect of category structure in our results is also interesting. Given the same amount of input, children's generalization in the 16-4 trials did not differ from that in the 16-2 trials. This contrasts with predictions from the hierarchical Bayesian model's instantiation of the Rational Hierarchical-inferential Approach (Kemp et al., 2007) that overhypothesis learning is more effective based on input consisting of more categories and fewer members per category, as opposed to the same amount of input consisting of fewer categories and more members per category. This proposed advantage is due to more categories providing more data for making higher-level inferences. It is likely that in our task, the difference between the 4-category (16-4 trials) and 2-category set (16-2 trials) was too subtle in influencing children's generalization. Future studies will explore a greater difference in category structure.

Questions remain as to the limits of children's overhypothesis learning, with respect to how domain- and modality-general it is, and how early in development it begins to emerge. Recently, Dewar and Xu (2010) show that 9-month-old infants can acquire second-order knowledge about how objects in a setting are organized into groups (by shape or color). This finding that even pre-verbal infants are capable of creating overhypotheses suggests that the mechanisms for overhypothesis learning are in place early on, and are domain-general, not limited to language learning. Future research can explore how overhypothesis formation goes beyond perceptual dimensions, and how it may be applied by children in learning in other domains.

This paper presents findings of how children can quickly and flexibly "learn to learn" in a linguistic task. These results are consistent with the idea that learners approach learning about something in the world on multiple levels. They hold implications for the origins of learning biases, in perhaps different cognitive domains, not just in language learning. Testing the generality and developmental origins of the ability exhibited in the current study is an important next step.

Acknowledgments

This work was supported by the McDonnell Foundation Collaborative Initiative on Causal Learning. We thank the UCB Infant Cognition and Language Lab members for their help in data collection, and the families for participating.

References

Baillargeon, R. (2008). Innate ideas revisited: For a principle of persistence in infants' physical reasoning. *Perspectives on Psychological Science, 3*, 2-13.

Carey, S., & Spelke, E. (1996). Science and core knowledge. *Philosophy of Science, 63*, 515–533.

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

Dewar, K., & Xu, F. (2010). Induction, overhypothesis, and the origin of abstract knowledge: Evidence from 9-month-old infants. *Psychological Science, 21*, 1871-1877.

Gergely, G., & Csibra, G. (2003). Teleological reasoning in infancy: the naive theoryof rational action. *Trends in Cognitive Science, 7 (7)*, 287–292.

Goodman, N. (1955). *Fact, fiction, and forecast.* Cambridge, MA: Harvard University Press.

Keil, F. C. (1981). Constraints on knowledge and cognitive development. *Psychological Review, 88,* 197-227.

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

Kemp, C., Goodman, N. D., & Tenenbaum, J. B. (2010). Learning to learn causal models. *Cognitive Science,* 1-59.

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

Lucas, C., Gopnik, A., & Griffiths, T. (2010). Developmental differences in learning the forms of causal relationships. *Proceddings of the Thirty-second Annual Conference of the Cognitive Science Society*.

Macario, J. F., Shipley, E. F., & Billman, D. O. (1990). Induction from a single instance: formation of a novel category. *Journal of Experimental Child Psychology, 50*, 179–199.

Mareschal, D., & Tan, S. H. (2007). Flexible and context-dependent categorization by eighteen-month-olds. *Child Development, 78*, 19-37.

Markman, E. M. (1989). *Categorization and Naming in Children*. Cambridge, MA: MIT Press.

Perfors, A., Tenenbaum, J. B., Griffiths, T., & Xu, F. (in press). A tutorial introduction to Bayesian modesl of cognitive development. *Cognition*.

Perfors, A., & Tenenbaum, J. B. (2009). Learning to learn categories. In N. Taatgen, H. van Rijn, L. Schomaker, & J. Nerbonne (Eds.), *Proceedings of the Thirty-first Annual Conference of the Cognitive Science Society* (p. 136-141).

Quine, W. V. O. (1960). *Word and Object*. Cambridge, MA: MIT Press.

Samuelson, L. (2002). Statistical regularities in vocabulary guide language acquisition in connectionist models and 15-20-month-olds. *Developmental Psychology*, *38*, 1016-1037.

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

Xu, F., Dewar, K., & Perfors, A. (2009). Induction, overhypotheses, and the shape bias: Some arguments and evidence for rational constructivism. In B. M. Hood and L. Santos (Eds.), *The Origins of Object Knowledge* (pp. 263-284). Oxford University Press.

Xu, F., & Tenenbaum, J. B. (2007). Sensitivity to sampling in Bayesian word learning. *Developmental Science, 10*, 288-297.