

What are you trying to tell me? A Bayesian model of how toddlers can simultaneously infer property extension and sampling processes

Hyowon Gweon, Joshua B. Tenenbaum, and Laura E. Schulz
(hyora, jbt, lschulz@mit.edu)

Department of Brain and Cognitive Sciences,
Massachusetts Institute of Technology
Cambridge, MA 02139 USA

Abstract

Young human learners possess a remarkable ability to make inductive inferences from sparse data. Recent research suggests that children's generalizations are sensitive to the process by which data are generated (i.e., teacher-driven vs. learner-driven sampling; Xu & Tenenbaum, 2007). In general, sampling process and properties of objects are tightly coupled; knowing how the data were sampled can inform your inference about property extensions, and vice versa. In real-world situations, however, both the extension of novel properties and the sampling process may be ambiguous. These situations commonly arise when children are learning socially from adults. How do children confront the challenge of simultaneously inferring both the property extension and the sampling process from a small amount of data? Here we present a Bayesian model showing how this joint inference problem can be solved. Consistent with the predictions of the model, two behavioral experiments suggest that toddlers (mean: 16 months) can use the relationship between a sample and a population to infer both the sampling process and the extent to which a non-obvious object property should be generalized.

Keywords: Sampling process; property induction; Bayesian model, cognitive development; toddlers

Previous developmental research suggests that young children have a remarkable ability to make inferences from small amounts of data. Children can learn object kinds, causal properties, and word meanings from just a few trials of evidence (Carey & Bartlett, 1978; Gopnik, Sobel, Schulz, & Glymour, 2001; Mandler & McDonough, 1996). Indeed, even infants can generalize a non-obvious property of an object (e.g., squeaking) to a similar-looking object after just a single exposure (Baldwin, Markman, & Melartin, 1993).

However, the process by which data are generated can influence the learner's inferences. Strikingly, even 8-month-olds are sensitive to sampling processes: they expect that a blind-folded draw from a bin of balls should result in a sample representative of the population (Xu & Garcia, 2008). Recent research suggests different sampling contexts are particularly important when learning in social contexts. For example, three- and four-year olds generalize a label for a novel object more conservatively when the exemplars are sampled by a knowledgeable teacher than when they are chosen by the learners themselves (Xu & Tenenbaum, 2007). Another recent study showed that children can even make sophisticated inferences about mental states of others given

a goal-directed, non-random sampling process. When an agent (i.e., a squirrel puppet) deliberately draws a non-representative sample of toys from the population, preschoolers infer that the agent has a preference for the sampled objects (Kushnir, Xu, & Wellman, 2008). Recent modeling work on preference learning shows that children can make these inferences by assuming that the agent selects the samples based on their subjective utility of each example (Lucas, Griffiths, Xu, & Fawcett, 2008). Note, however, that in these studies, the sampling process was specified by information about the person drawing the sample: the agent drawing the sample was either a knowledgeable adult or the child herself (Xu & Tenenbaum, 2007), was blindfolded or had visual access (Xu & Garcia, 2008), or was clearly intentionally selecting specific objects (Kushnir et al., 2008).

Frequently however, the sampling process may be less obvious; the learner may need to make inferences about both the sampling process and the properties of objects. Consider an example where mom pulls a few blue toys out of a box of blue and yellow toys and squeaks the blue toys. Did she select just the blue ones because only the blue toys squeak? Or did she sample from the whole box and just happened to pull out all blue toys that squeak? How can the child tell, and how is this inference about the sampling process linked to their inference about the property of the yellow toy?

Problems of this nature are much more difficult, because inferences about sampling process and property extension are tightly coupled: inferences about one should affect inferences about the other, but you need to infer both simultaneously. This chicken-and-egg problem comes up particularly in social contexts, where children are learning from other people. Because the cues to the appropriate sampling process may be ambiguous or too subtle for young learners to catch, they might not be able to use information about the sampling process to learn the extension of a demonstrated property from a few samples of evidence.

This challenge poses an interesting question: how do young children confront a situation in which both variables need to be inferred? Here we provide a simple Bayesian model that captures how this joint inference problem can be solved. We then present the first empirical evidence that toddlers (mean: 16 months) can engage in these joint inferences to determine both the nature of the sampling process and the extension of non-obvious object properties.

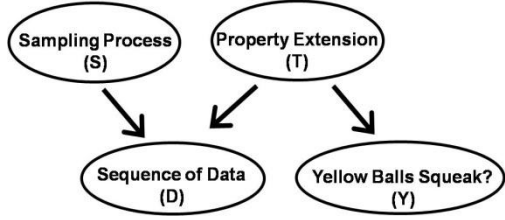


Figure 1: A simple Bayes net describing the joint inference problem in the current study.

We end with a discussion of the relationship between the model predictions and behavioral results, and its implications for the mechanisms that support inductive generalization in early childhood.

A Bayesian Model

Our example forms the basis for our experiment. Children saw an experimenter pull either one or three blue balls from a box filled with blue and yellow balls. They were shown that the blue ball(s) squeaked when squeezed. The question was whether children would expect this property to generalize to the yellow ball. We measured these expectations based on the toddler's exploratory behavior: after observing the data of one or more squeaking balls, children were given a yellow ball, which did not in fact squeak. We recorded the number of times the children tried to squeeze the yellow ball. The more times they squeeze, the more we can attribute to them the expectation that the yellow balls squeak. We varied the number of balls observed, as well as the proportion of blue and yellow balls in the box. These factors determine how likely the observed sample appears to be under different joint hypotheses about the sampling process and the property's extension, and we will study how children's inferences in different situations relate to an ideal Bayesian analysis.

The joint inference problem facing children in this situation can be described in terms of a simple Bayes net (Figure 1). The learner observes data $D = n$ examples of blue balls that squeak, which are drawn from a box containing a fraction (β) of blue balls and $1-\beta$ yellow balls. Because the task does not provide explicit cues about the sampling process, the learner needs to make inferences about both the sampling process S and the property extension T purely from data D . For simplicity, we consider only three possible values for T (t_1 : the property applies only to blue balls; t_2 : only to yellow balls; t_3 : to all balls) and two possible values for S (s_1 : randomly sampling from just the squeaking set of balls, specified by T ; s_2 : randomly sampling from the whole box).

The learner's goal is to predict Y , the proposition that yellow balls squeak. Y depends directly on T , not S or D ; given that we know the set of balls that squeak, the observed data or the process by which the data were sampled is irrelevant to predicting whether the yellow balls squeak. However, inferences about T from D must take into account the different possible values of S ; formally, our Bayesian

analysis must integrate out S in scoring each value of T . Because the data are inconsistent with hypothesis t_2 , only two hypotheses for T are relevant; t_3 predicts that yellow balls squeak while t_1 predicts that they do not. Following Tenenbaum and Griffiths (2001), the evidence for one of these hypotheses over the other can be measured by the likelihood ratio:

$$L = \frac{P(D|t_3)}{P(D|t_1)} = \frac{P(n|t_3, \beta)}{P(n|t_1, \beta)}$$

We posit that children's exploratory behavior -- how much they squeeze the yellow ball, expecting a squeak -- will be monotonically related to L . This analysis makes predictions that are independent of the prior probabilities children assign to t_1 or t_3 , removing a degree of freedom that would otherwise need to be measured or fit empirically to their behavior.

These likelihoods can be computed by integrating out the sampling process:

$$P(n|t, \beta) = \sum_{s_i \in S} P(n|t, s, \beta)P(s).$$

To evaluate these likelihoods we need the following four conditional probabilities:

$$\begin{aligned} P(n|t_1, s_1, \beta) &= 1. \\ P(n|t_1, s_2, \beta) &= \beta^n. \\ P(n|t_3, s_1, \beta) &= \beta^n. \\ P(n|t_3, s_2, \beta) &= \beta^n. \end{aligned}$$

Let α denote the prior probability $P(s_1)$, that the experimenter is sampling from just the squeaky balls; $P(s_2) = 1 - \alpha$. We then have:

$$\begin{aligned} P(n|t_1, \beta) &= \sum_{s_i \in S} P(n|t_1, s, \beta)P(s) \\ &= P(n|t_1, s_1, \beta)P(s_1) + P(n|t_1, s_2, \beta)P(s_2) \\ &= \alpha + \beta^n(1 - \alpha). \end{aligned}$$

$$\begin{aligned} P(n|t_3, \beta) &= \sum_{s_i \in S} P(n|t_3, s, \beta)P(s) \\ &= P(n|t_3, s_1, \beta)P(s_1) + P(n|t_3, s_2, \beta)P(s_2) \\ &= \beta^n\alpha + \beta^n(1 - \alpha) \\ &= \beta^n. \end{aligned}$$

The likelihood ratio, measuring the evidence in favor of the proposition that yellow balls squeak, is then:

$$\begin{aligned} L &= \frac{P(n|t_3, \beta)}{P(n|t_1, \beta)} \\ &= \frac{\beta^n}{\beta^n(1 - \alpha) + \alpha}. \end{aligned}$$

Experiment 1

In this behavioral experiment we asked whether toddlers could solve this joint inference problem. Children saw an experimenter draw blue balls from a box and were then given the yellow ball to play with. We varied the parameters n and β to provide: 1) evidence that could have been generated in either sampling processes (Blue_3balls condition: $n = 3$, $\beta = 0.75$), or 2) evidence for a very suspicious coincidence that might suggest a non-random sampling process (Yellow_3balls condition: $n = 3$, $\beta = 0.25$). As we show in the results section, our Bayesian analysis predicts very different strengths of evidence in these two cases.

Methods

Participants Thirty toddlers (mean: 15 months, 24 days; range: 13 to 18 months) were recruited from a local children’s museum, and randomly assigned to Blue_3balls condition or Yellow_3balls condition. Children were excluded from analysis if they did not complete the procedure due to fussing out or if they did not interact with the target object at all. Three children were excluded and replaced for these reasons.

Materials Two foam-board boxes were constructed (30 x 45cm x 30 cm). Each box had a hidden compartment in the back. One box contained 12 blue balls and 4 yellow balls (henceforth the Blue Box), and the other contained 4 blue balls and 12 yellow balls (henceforth the Yellow Box). The front side of the boxes was transparent, and all 16 balls were visible through the transparent window. The blue balls had a squeaking mechanism inside. The squeaking mechanism was removed from the yellow balls so that they were inert. Additionally, the yellow balls had a wooden handle with a bell-shaped object at the end (providing an additional ‘banging’ affordance so the child could readily engage in a behavior other than squeezing the balls). Thus the objects were perceptually similar (an adult would categorize them all as ‘dog toys’) but not identical. The boxes had a small opening at the top, allowing the experimenter to pull out the balls from the hidden compartment. Thus even when the balls were pulled from the box, the view from the front of the box (showing all 16 balls) stayed constant.

Procedure Children were tested individually in a quiet lab room at the Children’s Museum. The child sat on a highchair or on a small stool; the parent sat behind the child, out of the child’s line of sight. A box (Blue or Yellow box depending on condition) sat on a low table in front of the child. Children saw the Blue Box (blue:yellow = 3:1) in Blue_3balls condition, and the Yellow Box (blue:yellow = 1:3) in Yellow_3balls condition. The experimenter drew the child’s attention by pointing to the transparent window and the contents of the box. In all conditions, the experimenter had informational access to the contents of the box. However, her action was identical across conditions: there was no way to tell whether she was sampling from a specific

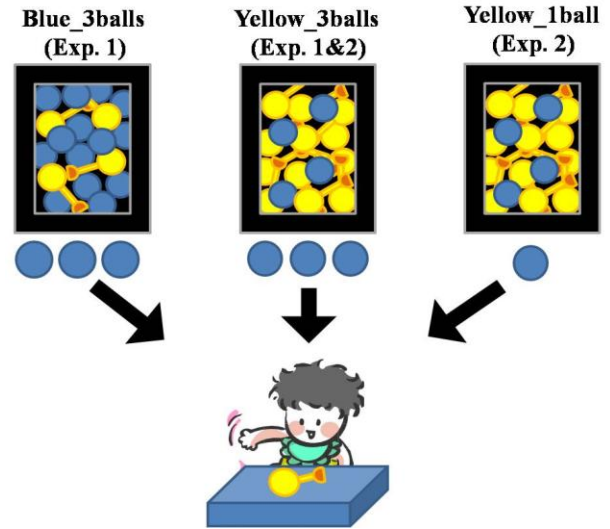


Figure 2: A schematic drawing of the experimental procedures in Experiments 1 and 2.

subset of balls in the box or sampling from all objects in the box. Each time she pulled out a ball, she squeezed the ball so that it squeaked and then set that ball on the table. She repeated this procedure to pull out a total of three blue balls in both conditions. The experimenter then pulled out a yellow ball and put it in front of the child. The child was allowed to play freely with the ball for 30 seconds.

Results & Discussion

We coded both the number of children who squeezed the yellow ball and the number of times each child tried to squeeze the ball during the 30 seconds of free play. The data was coded either by the first author or a second coder. An additional coder, blind to the experimental conditions recoded all of the data. Inter-coder reliability averaged 94%.

Under the Bayesian framework described earlier, children might consider the following four joint hypotheses about the sampling process and property extension:

- H1: sampling = squeaking set (s_1), property = blue (t_1)
- H2: sampling = whole box (s_2), property = blue (t_1)
- H3: sampling = squeaking set (s_1), property = all (t_3)
- H4: sampling = whole box (s_2), property = all (t_3).

In Yellow_3balls condition, when three blue balls ($n = 3$) are pulled and squeaked from a box that contains only $\frac{1}{4}$ blue balls ($\beta = 0.25$), the sample is very unlikely under the random sampling assumption; the data support s_1 . This inference is tightly coupled to the property extension. If the experimenter is selecting the balls from the squeaking set only, the data are most consistent with the hypothesis that just the blue balls have the squeaking property. Therefore, the joint hypothesis H1 makes the observed sequence of data more likely than any of the three other alternatives. In Blue_3balls condition, however, the data do not discriminate between s_1 and s_2 : a random sampling process

could generate three blue balls from the Blue Box. If the sampling process could have been random, the data also do not provide evidence that only the blue balls squeak, as opposed to all the balls. Therefore, children should attempt to squeak the yellow ball less often in Yellow_3balls condition than in Blue_3balls condition.

The experimental results confirmed our prediction: mean number of squeezes was significantly fewer in Yellow_3balls than in Blue_3balls condition (2.53 vs. 0.87; $t(28) = 2.45$, $p < 0.05$, see Figure 3). Consistent with this result, fewer children squeezed the ball in Yellow_3balls than in Blue_3balls condition (33% vs. 80%; $\chi^2(1, N=30) = 8.89$, $p < .005$). These results suggest that toddlers' inferences about the sampling process and property extension led them to constrain their generalization of the squeaking property to the blue ball in Yellow_3balls condition, whereas children in Blue_3balls condition were willing to generalize the property to both balls. This is consistent with our model predictions: assuming that two sampling hypotheses (s_1 and s_2) are equal a priori ($\alpha = 0.5$), the likelihood ratio $L = \frac{P(n|t_3, \beta)}{P(n|t_1, \beta)}$ is 0.59 for Blue_3balls ($n=3$, $\beta = 0.75$), and 0.03 for Yellow_3balls condition ($n=3$, $\beta = 0.25$) (see Figure 3).

Experiment 2

Experiment 1 suggests that toddlers can infer both the sampling process and extension of a novel object property from data, even when the sampling process is otherwise ambiguous. The results are consistent with the predictions provided by our Bayesian analysis. However, it is possible that the children were simply sensitive to the ratio of objects in the box. Note that the blue balls were the majority of objects in the box in Blue_3balls condition (blue:yellow = 3:1) and the minority in Yellow_3balls condition (blue:yellow = 1:3). If children assumed that properties of the majority-object could be generalized to the minority-object, they would generalize the squeaking property to the yellow ball in Blue_3balls condition and not in Yellow_3balls condition.

In Experiment 2, we addressed this alternative explanation by running a condition (Yellow_1ball condition) in which we draw just one blue ball out of the Yellow Box. Even though $\frac{3}{4}$ of the balls in the box are yellow, a single blue ball could be drawn from this box as a result of a random sampling process. This single piece of data does not discriminate between s_1 and s_2 and it also does not provide strong evidence for either t_1 or t_3 . Because we wanted to replicate our results internally and ensure that the children really would constrain their squeezing when the sample was non-representative, we also ran another condition in which three blue balls were drawn from the Yellow box (Yellow_3balls replication, identical to Yellow_3balls condition in Experiment 1). The prediction is that the children restrict their generalization of the squeaking property to the blue ball significantly more in Yellow_3balls replication than in Yellow_1ball condition (i.e., the results

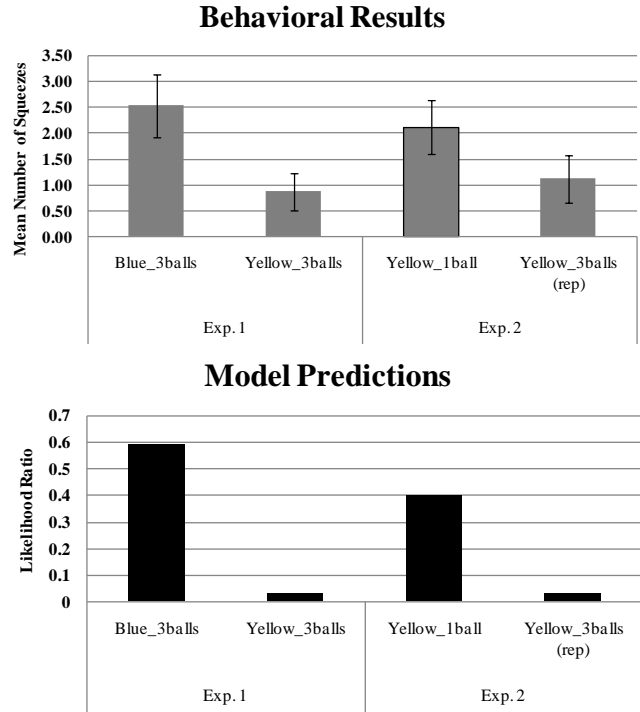


Figure 3: The Likelihood ratios predicted by the model (top) and the mean number of squeezes from Experiments 1 & 2 (bottom).

of Yellow_3balls replication should replicate Yellow_3balls condition in Experiment 1, while the results of Yellow_1ball condition should mirror Blue_3balls condition in Experiment 1).

Methods

Participants Thirty-four toddlers (mean: 15 months, 14 days; range: 13 to 18 months) were recruited from a local children's museum and randomly assigned to Yellow_1ball condition or Yellow_3balls replication. The same criteria as in Experiment 1 were used to exclude children from analysis: seven toddlers were excluded and replaced for these reasons. Two additional children were excluded and replaced, one due to experimental error and one due to parental interference.

Materials The same materials as in Experiment 1 were used.

Procedure The experimental procedure was identical to that of Yellow_3balls condition in Experiment 1 except that in Yellow_1ball condition, only one blue ball was drawn from the box. Yellow_3balls replication was an exact replication of Yellow_3balls condition.

Results & Discussion

The results were coded as in Experiment 1. The results were consistent with our predictions: there was a trend for children to squeeze less often in Yellow_3balls replication

than in Yellow_1ball condition (2.12 vs. 1.12; $t(32) = 1.48$, $p = 0.14$, see Figure 3)¹ and significantly fewer children squeezed the ball in Yellow_3balls replication than in Yellow_1ball condition (41% vs 82%; $\chi^2(1, N=34) = 6.1$, $p < .05$). These results were also consistent with our model predictions. Assuming $\alpha = 0.5$ as in Experiment 1, the likelihood ratio $L = \frac{P(n|t_3, \beta)}{P(n|t_1, \beta)}$ is 0.40 for Yellow_1ball condition ($n=1$, $\beta = 0.25$) and 0.03 for Yellow_3balls replication ($n=3$, $\beta = 0.25$) (see Figure 3). Also as predicted, the results of Yellow_1ball condition mirrored those of Blue_3balls condition (number of children squeezing: 82% vs. 80%, $p = ns$; mean squeezes: 2.53 vs. 2.12, $p = ns$) and Yellow_3balls replication replicated Yellow_3balls condition (number of children squeezing: 33% vs. 41%, $p = ns$; mean squeezes: .83 vs. 1.12, $p = ns$).

Thus although blue balls were the minority object in both conditions of Experiment 2, children were willing to generalize the property to the yellow ball when the blue ball was potentially randomly sampled from the box (Yellow_1ball condition). When the sample was unlikely to be randomly generated (Yellow_3balls replication), children constrained their generalization. Note further that the toddlers squeaked the ball more often in Yellow_1ball condition than in Yellow_3balls replication, even though they actually saw a ball being squeaked more often in Yellow_3balls replication (three times) than in Yellow_1ball condition (once). Therefore, children's tendency to squeak the ball was unrelated to the number of times they actually saw the target action but was well predicted as a joint inference about the sampling process and the property extension.

General Discussion

In the current study we presented a formal Bayesian account that captures how the sampling process and extension of object properties can be simultaneously inferred from a small sample of data. We also provided evidence that 16-month-olds can correctly solve this joint inference problem in the way that is consistent with the model's predictions.

In learning a concept, object label, or an object property from sampled examples, there are at least two possible sampling processes that give rise to an observed set of positive examples. To decide whether to generalize an object label or property to beyond the examples given, it is important to know how the examples are generated. In the context of the current study, the hypothesis that the experimenter is sampling from just the squeaking set of balls (s_2) represents a kind of 'pedagogical' sampling or 'preference-based' sampling: the experimenter knows which balls squeak, and is deliberately choosing just those balls

either because she wants to show them to the child or because she likes them. Tenenbaum and Griffiths (2001) referred to this mode of sampling as 'strong sampling' in which the examples are sampled from just the property's extension, as opposed to 'weak sampling' where the exemplars are drawn from all available objects independently of their properties and the sampled objects happen to have that property by chance.

Previous work by Xu and Tenenbaum (2007) shows that 3- and 4-year-olds can infer which sampling mode is more appropriate given the pragmatics of the learning situation. In this study, children were given a set of objects spatially sorted according to their basic and subordinate categories. When a knowledgeable speaker provided three examples of objects from the same subordinate category and paired them with a label by saying "this is a blicket" (teacher-driven condition), children restricted their generalization of this label to a very specific set of objects that are similar to the exemplars. However, when only the first example was given by the teacher and the two additional examples were drawn by the children themselves from the same subordinate category (learner-driven condition), they generalized the label more broadly at the basic category level. These results suggest that children readily recognize the difference in sampling processes and apply the appropriate sampling assumption in order to support their generalization.

The authors provide a Bayesian analysis to explain their results. In the teacher-driven condition, an example of three objects from the same subordinate category is a highly suspicious coincidence had the teacher been sampling from the whole set of objects. Under strong sampling assumption, the *size principle* (Tenenbaum & Griffiths, 2001) strongly favors a smaller set of objects (i.e., the same subordinate category) as the population from which the teacher was sampling from. In the learner-driven condition, however, only the first example was given by strong sampling and the subsequent examples were drawn by the children under weak sampling. Although the learner observes the same object-label pairings as in the teacher-driven condition, the likelihoods of observing the data given different hypotheses about the concept extension do not provide strong evidence for subordinate-level interpretation of the label over the basic-level interpretation.

Note that, however, the evidence for the appropriate sampling context was given by the pragmatics of the task in this study: a knowledgeable adult was sampling in the teacher-driven condition, and children were implicitly encouraged to choose examples just from the same subordinate category as the first example in the learner-driven condition. In real-world learning situations, the process by which the data is generated may not be as obvious: for example, the behavior of the agent may fail to distinguish sampling from a subset of the population versus sampling from a whole population (as in the current experiment) or the social cues that might indicate the sampling process may be too subtle for young learners who may not yet have a fully-fledged theory of mind (Gopnik &

¹ One child in Yellow_3balls replication was an outlier - squeezing the ball 3 standard deviations more than the mean (7 vs. 1.12 squeezes). Excluding that child from the analyses, the difference in the mean number of squeezes between conditions C and D is then also significant (2.12 vs. 0.75; $t(31) = 2.35$, $p < 0.05$).

Astington, 1988; Wellman, Cross, & Watson, 2001; Wimmer & Perner, 1983). The current study shows for the first time that children as young as 16 months old can infer the sampling process purely from observing the sample.

Our model suggests that children use the number of samples(n) and the proportion of blue and yellow balls in the population(β) to make joint inferences about the sampling process and property extension. In future work, one might parametrically vary n and β to see whether children's responses vary accordingly, or present different data sequences that might suggest other sampling processes (thus varying α , which was set to 0.5 in our model).

A related but alternative possibility is that children solve the problems in our study, not as a simultaneous inference, but by first inferring the sampling process and then using this information to infer the property extension. Because they infer that the experimenter *intentionally* selected three blue balls to show that they squeak, they might conclude that just the blue balls have the property. Although the underlying processes are similar to what our model has proposed, future work might distinguish between these possibilities.

We found that the mean number of squeezes across conditions was consistent with the model predictions. However, it is still unclear how exactly this behavioral measure correlates with the likelihood ratios from the model. Here we assumed that the mean number of squeezes might reflect the strength of children's belief that the yellow ball has the squeaking property: because the yellow ball was inert, children might perseverate to the degree that they are convinced that the yellow ball should squeak. The likelihood ratios reflect the strength of evidence that the sample provides for discriminating the two hypotheses about property extension (all balls squeak vs. only blue balls squeak). Although it is reasonable to assume that these two measures might be highly correlated, note that the differences between the group means in the number of squeezes were mainly driven by the children who did not squeeze at all. Note, further, that the all-or-none measure of whether a child squeezed or not showed the same qualitative pattern as the mean number of squeezes. Further research in both computational modeling and behavioral experiments should aim to clarify what aspects of the behavior the model is predicting.

Our findings bear on the theoretical stance that humans are rational learners from very early in development. Inferring both the property extension and sampling process is one of the chicken-and-egg problems that children frequently encounter in real-world learning situations. Our study suggests that young learners possess a powerful mechanism to support inductive inference in early childhood. Through this mechanism, children can not only make generalizations from sparse data but can also, even in the absence of explicit social cues, do so by inferring and integrating the processes by which the data are generated.

Acknowledgments

Many thanks to Phoebe Neel for help with data collection, to Sydney Katz for help with blind coding, and to members of the Early Childhood Cognition Lab for helpful comments and suggestions. This research was supported by an NSF Faculty Early Career Development Award and a John Templeton Foundation Award to L.S. and a James S. McDonnell Foundation Collaborative Interdisciplinary Grant on Causal Reasoning to LS and JT.

References

- Carey, S., & Bartlett, E. (1978). Acquiring a single word. *Papers and Reports on Child Language Development*, 15, 17-29.
- Gopnik, A., & Astington, J. W. (1988). Children's understanding of representational change and its relation to the understanding of false belief and the appearance-reality distinction. *Child Development*, 59(1), 26-37.
- Gopnik, A., Sobel, D. M., Schulz, L. E., & Glymour, C. (2001). Causal learning mechanisms in very young children: two-, three-, and four-year-olds infer causal relations from patterns of variation and covariation. *Developmental Psychology*, 37(5), 620-629.
- Kushnir, T., Xu, F., & Wellman, H. M. (2008). Preschoolers use statistical sampling information to infer the preferences of others. In V. Sloutsky, B. Love, & K. MacRae (Eds.) *Proceedings of the 30th Annual Conference of the Cognitive Science Society.*, 1563-1566.
- Lucas, C., Griffiths, T., Xu, F., & Fawcett, C. (2008). A rational model of preference learning and choice prediction by children. *Proceedings of the 22nd Annual Conference on Neural Information Processing Systems.*
- Mandler, J. M., & McDonough, L. (1996). Drinking and driving don't mix: inductive generalization in infancy. *Cognition: International Journal of Cognitive Science*, 59(3), 307-335.
- Tenenbaum, J. B., & Griffiths, T. L. (2001). Generalization, similarity, and Bayesian inference. *Behavioral and Brain Sciences*, 24(4), 629-640.
- Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: the truth about false belief. *Child Development*, 72(3), 655-684.
- Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition: International Journal of Cognitive Science*, 13(1), 103-128.
- Xu, F., & Garcia, V. (2008). Intuitive statistics by 8-month-old infants. *Proceedings of the National Academy of Sciences*, 105(13), 5012-5015.
- Xu, F., & Tenenbaum, J. B. (2007). Sensitivity to sampling in Bayesian word learning. *Developmental Science*, 10(3), 288-297.