The emergence of probabilistic reasoning in very young infants

Stephanie Denison (smdeniso@berkeley.edu), Christie Reed (cereed@berkeley.edu)
& Fei Xu (fei_xu@berkeley.edu)
Department of Psychology, University of California, Berkeley, CA 94720 USA

Abstract

How do people make such rich inferences from such sparse data? Recent research has explored this inferential ability by investigating probabilistic reasoning in infancy. For example, 8- and 11-month-old infants can make inferences from samples to populations and vice versa (Denison & Xu, 2010a; Xu & Garcia, 2008). The current experiment investigates the developmental origins of this probabilistic inference mechanism with 4- and 6-month-old infants. Infants were shown 2 large boxes, 1 containing a ratio of 4 pink to 1 yellow balls, the other containing the opposite ratio. The experimenter sampled from, e.g., the mostly pink box, and removed either 4 pink and 1 yellow or 4 yellow and 1 pink ball on alternating trials. 6- but not 4-month-olds looked longer at the 4 yellow and 1 pink ball sample (the improbable outcome) than the 1 yellow and 4 pink ball sample (the probable outcome).

Keywords: Statistical inference; Probabilistic reasoning; Infant cognition.

Introduction

A wide range of evidence exists to show that human learners are capable of making large inductive leaps in the face of very small amounts of data. It is easy to appreciate why this might be such an important and ubiquitous ability. For example, imagine a person from our hunter-gatherer ancestry trying to determine which types of trees produce berries that are good for eating. Let's say they sample roughly five berries from a couple of trees and find that one tree produces four good tasting berries and the other only produces one or two. They may make the inference that all berries from the former tree are good for eating and that the latter tree type should be avoided.

What is the cognitive mechanism that allows human learners to make such rapid and often highly accurate inductive inferences with such small amounts of data? Recent literature in cognitive development has focused on the origins of statistical inference in infancy as a possible starting point. At least three lines of work have emerged exploring various aspects of infant probabilistic reasoning: First, we know that infants as young as 8 months are capable of making inferences from small samples to large populations as well as from populations to samples (Xu & Garcia, 2008). Second, we know that 12-month-old infants can make inferences from populations to samples when reasoning about single-event probability (Denison & Xu, 2010b; Teglas, Girotto, Gonzalez & Bonatti, 2007). Third, evidence suggests that infants as young as 11 months take into account the implications of sampling conditions (e.g., random vs. non-random sampling) and object properties (e.g., solidity and cohesion) when making these inferences (Denison & Xu, 2010a; Gweon, Tenenbaum & Schulz, 2010; Teglas et al. 2007; Xu & Denison, 2009).

In the current experiment, we explore the age at which infants begin to make inferences from samples to populations. We ask whether 4- and 6-month-old infants can make basic probabilistic inferences using a variant of the paradigm first introduced by Xu and Garcia (2008). In their experiments, a violation of expectancy looking-time paradigm was employed to reveal whether 8-month-old infants have an intuitive ability to make generalizations from samples to populations. In Experiment 1, infants were shown samples being drawn from a large covered box and, on alternating trials the experimenter either removed 4 red balls and 1 white ball or 4 white balls and 1 red ball. Then the experimenter revealed the population of balls in the box – a 9:1 ratio of red to white balls. Eight-month-old infants looked longer at the 4 white and 1 red ball sample (the improbable outcome) than the 4 red and 1 white ball sample (the probable outcome; see Figure 1).

Although at first blush this appears to suggest that infants have a rudimentary ability to reason about probability, the authors note that two interpretations of this looking-time pattern exist: The first, which we will call the "probabilistic account", suggests that infants looked longer at the 4 white and 1 red ball sample because they understand the predictive relationship between samples and populations and thus they considered it to be a relatively improbable sample. The second, termed here the "perceptual mismatch account", suggests that infants simply prefer to look at displays in which the population box and sample container contrast in perceptual appearance. That is, infants simply looked longer at trials displaying the less probable sample because it created a perceptual mismatch between the two displays present on stage (see the outcomes in Figure 1). This account represents a lower-level interpretation of infant performance, as it predicts an identical looking pattern as the probabilistic account but does not require that infants understand anything about the relationship between the sample and population.[1]

To distinguish among these accounts, Xu and Garcia (2008, Expt. 3) designed an experiment in which the 4:1 sample of balls was no longer drawn from the population box. Another group of 8-month-old infants participated in a procedure that was equivalent to the one just described except that the relationship between the sample and

[1] Adults viewed the Expt. 1 displays and rated the improbable outcome as "unexpected" and the probable outcome as "expected." They did not note perceptual mismatches or probability in their explanations. This suggests that computations of probabilities may generally be largely implicit and inaccessible to conscious thought.

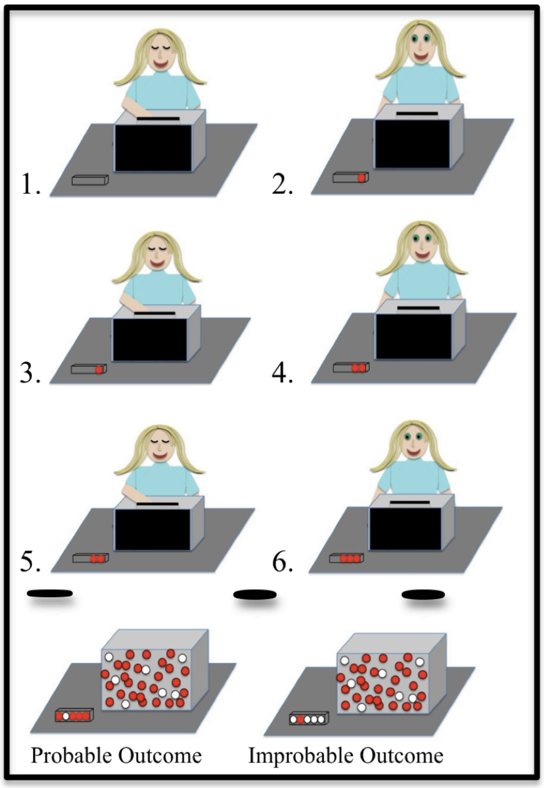

Figure 1: The sequence of a test trial in Expt. 1 (Xu & Garcia, 2008). The experimenter shakes the box, closes her eyes and draws out balls from the closed box. She then reveals the population.

population was eliminated. Test trials began in the exact same manner as Experiment 1 with the closed large box and empty sample container present on the stage. During the sampling phase, the experimenter shook the box but instead of reaching into the box to draw the sample of either 4 white and 1 red ball or 4 red and 1 white ball, she reached into her pocket to draw the sample instead. She then revealed the contents of the box just as in Experiment 1. This resulted in identical test trial displays to those in Experiment 1 but in this case, infants should have no reason to expect a relationship between what they saw in the small container and the population box presented on the stage. The 8-month-old infants looked about equally when the mostly red box was displayed with the 4 red and 1 white ball sample (the perceptual match) or the 4 white and 1 red ball sample (the perceptual mismatch). This suggests that neither display violated infants' expectations when the relationship between the box and container was eliminated. This result provides evidence in favor of the probabilistic account of infants' performance in Experiment 1, i.e., that infants were in fact reacting to the relative improbability of the sample and not the perceptual mismatch between the box and sample.

In a follow-up experiment, we began to explore whether even younger infants have some of these probabilistic intuitions in place. It is plausible that younger infants could succeed at a version of this task, given that evidence exists revealing statistical learning in 2- to 6-month-old infants in domains such as rule learning, phoneme discrimination, and visual pattern learning (e.g., Dawson & Gerken, 2009; Kirkham, Slemmer & Johnson, 2002; Maye, Werker, & Gerken, 2002).

We tested a group of 6-month-old infants using the same procedure of Xu and Garcia (2008). The findings were inconclusive. Infants performed as expected in the replication of Experiment 1, looking longer at trials in which the experimenter sampled 4 white and 1 red ball than 4 red and 1 white ball from the mostly red population. However, infants continued to follow this looking pattern in the control experiment during which the experimenter drew from her pocket (i.e., the replication of Expt. 3).

Although this pattern of findings does not support the probabilistic account, it also does not necessarily rule out the possibility that infants did make a correct generalization from sample to population in the Experiment 1 replication. It is possible that 6-month-olds do appreciate the relationship between samples and populations but they also looked longer at the perceptual mismatch in the control task because they continue to react to the mismatch when the sample was drawn from the experimenter's pocket. Unfortunately the experimental design cannot tease apart the two interpretations when infants continue to look longer at the mismatches when the sample is not drawn from the box.

In the current study, we use an experimental design appropriate for testing younger infants in a task where the perceptual mismatch is eliminated but the displays remain easy to process. We equated the overall quantity of each ball color present on the stage in population boxes during test trials by keeping two complementary boxes on display throughout all trials (see Figure 2). After familiarizing infants with the two population boxes (one mostly pink; one mostly yellow), each test trial began with the two covered population boxes on the stage, one on the left and one on the right and a small transparent container to hold a sample in the middle. The experimenter drew the infants' attention to each of these boxes and drew a sample of, e.g., 4 pink and 1 yellow ball from the box on the right and placed it in the container. She then mimicked this action with the box on the left to equate the amount of attention paid to each box. Finally the experimenter revealed to the infant that the box on the right side of the stage had a 4:1 ratio of pink to yellow balls, and the box on the left side had the opposite ratio. On each trial the sample alternated between a 4 pink and 1 yellow sample (the more probable sample) and a 4 yellow and 1 pink sample (the less probable sample). If infants are only sensitive to perceptual mismatches and not sampling, they should look equally at all test trials, as the large boxes on display have equal amounts of each color and the sample therefore creates a slight but equal mismatch across every trial. If, on the other hand, infants are sensitive to the relationship between the sample and population, they will look longer on trials where the less probable sample is drawn from the relevant population box (e.g., the 4 pink and 1 yellow sample drawn from the mostly yellow box). We tested 4- and 6-month-old infants in this new design.

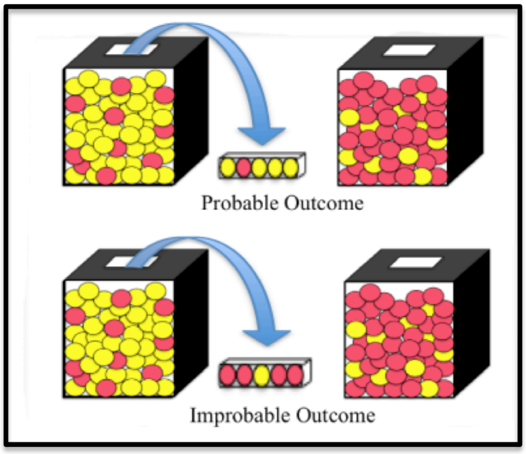

Figure 2: The two possible outcomes in the current experiment. The population boxes displayed simultaneously ensure equal amounts of pink and yellow are displayed.

Method

Participants

Participants were 32 infants: Sixteen 6-month-olds (6 males; *M* = 6;4 [months; days], *R*=5;15 to 6;17) and sixteen 4–month–olds (12 males; *M* = 4;15; *R* = 4;1 to 5;0). Ten infants (six 6-month-olds and four 4-month-olds) were tested but excluded due to fussiness (4), not attending during sampling (2), providing looking times over 3.5 standard deviations above the average (1) or parental interference (3). Infants were recruited from the San Francisco Bay Area and they received a small gift for participating in the study.

Materials

Ping-Pong Balls A total of 166 (83 yellow and 83 pink) ping-pong balls were used in the experiment.

Boxes and Containers A small open white box (15.5 cm x 17.5 cm x 8 cm) constructed from foam core was used in the "Free Play" phase of the experiment (see Procedure section). The box contained 3 yellow and 3 pink balls.

A small, transparent Plexiglas container with an open top (20 cm x 4.5 cm x 4.5 cm) was used to display the 5-ball samples during test trials. It was narrow enough that the balls lined up in a row when placed in the container.

Two large 31 cm x 23.5 cm x 23.5 cm boxes were used to display the populations during the familiarization and test trials. The boxes were black rectangular cubes and the inside of each box was partitioned into two parts with Plexiglas inserted in the front to provide a transparent window and a board in the middle to hold the populations of balls. A hidden center compartment measuring 10 cm x 24 cm x 10 cm was cut in the top of the two boxes to hold the samples to be removed from the box during test trials. From the infants' perspectives, the box appeared as one single unit, filled completely with ping-pong balls. The Plexiglas display windows were covered with a black fabric curtain that could be lifted to reveal the contents of the boxes through the Plexiglas windows. The "mostly pink" box contained 60 pink and 15 yellow ping-pong balls (pink to yellow = 4:1); the "mostly yellow" box contained the opposite (pink to yellow = 1:4). The boxes looked identical when viewed with the curtains down.

Apparatus

Testing occurred in a room divided in half by curtains spanning the width and height of the room and objects were presented on a puppet stage. The viewable area of the stage measured 94 cm x 55 cm (width x height). The experimenter sat behind the stage with her upper body and head visible to the infant. An observer, present to code the infant's looking times, sat in a corner of the room and was not visible to the infant. She watched the infant on a TV monitor and coded the infant's looking behavior online using JHAB version 1.0.0 (R. Casstevens, 2007). She was blind to the order of the test trials.

The infant sat in a high chair approximately 70 cm from the center of the stage. The parent sat next to the infant facing the opposite direction of the stage and was instructed to avoid looking at the stage. Two camcorders were used to record the experimental session, one to record the infant's looking behavior and the other to record the experimenter's presentation of the objects.

Design and Procedure

Calibration To calibrate each individual infant's looking window, a squeaky toy or keys were used to direct the infant's attention to the outside parameters of the stage.

Free Play Phase After calibration, the infant was shown the box with 3 pink and 3 yellow ping-pong balls. He/she was encouraged to play with the balls for approximately 30 seconds to allow infants to become familiar with the objects.

Familiarization trials (4 trials) To begin each trial, the experimenter placed the two large boxes on the stage 30 cm apart with the front curtains down. She shook the box on the right side of the stage a few times, saying, "What's in this box?" She then shook the box on the left a few times, saying, "What's in this box?" She lifted the front covers of both boxes simultaneously, revealing the separate populations of mostly pink and mostly yellow balls, and said "Look, [baby's name], look!" She put her head down and directed her gaze to the floor. The observer began timing upon hearing the second, "look". Trials ended when the infant looked away for 2 consecutive seconds.

The same two boxes of mostly pink and mostly yellow ping-pong balls were presented in the same locations for all 4 familiarization trials within a single experimental session. Between trials the boxes were removed from the stage and a black curtain was lowered at the back of the stage to conceal the experimenter. These trials were included to familiarize infants to the large boxes, as well as to the general procedure of the study. Additionally, exposing the infants to these two contrasting populations, one of mostly pink balls, the other of mostly yellow balls, might cue them to attend to ratios. The familiarizations lasted approximately 3 minutes.

Test trials (6 trials) On each test trial, the experimenter placed the two large boxes on the stage (keeping them in this same location on all 6 trials) with the front curtains closed. The experimenter always sampled from the box on her right. She shook each box one at a time just as in the familiarization trials while saying, "What's in this box?" She then closed her eyes, turned her head away, and reached into the box on her right. She pulled out 3 ping-pong balls and placed them into the small Plexiglas container in the middle of the stage one at a time. She then closed her eyes, turned her head away, and reached into the box on her left (not pulling out any ping-pong balls) and placed her hand on top of the small Plexiglas container in the middle of the stage to mimic the sampling motions made from the box on the right. She then repeated this action, pulling out 2 more ping-pong balls from the right hand box and placing them in the Plexiglas container and again mimicking this action with the box on the left. Thus the small Plexiglas container held a total of 5 balls. On alternating trials the sample removed from the population box was either 4 pink and 1 yellow balls or 4 yellow and 1 pink balls. Then the experimenter lifted the front covers of both boxes on the stage simultaneously and said "Look, [baby's name], look!" She put her head down and directed her gaze toward the floor. The observer began timing upon hearing the second, "look", and ended the trial after the infant looked away for 2 consecutive seconds. Between trials, the stage was cleared and the back curtain was lowered. The test trials lasted approximately 6 minutes.

Design The side that the population boxes (mostly pink or mostly yellow) were on and whether the infant saw the 4 pink and 1 yellow ball sample first were fully counterbalanced across infants.

Predictions

If infants are sensitive to the relationship between samples and populations (i.e., assuming random sampling, the composition of a sample is likely to reflect the overall composition of a population) they should look longer at test trials displaying outcomes that violate this expectation than outcomes that are in line with this expectation. Therefore infants who saw the experimenter sampling from the population with a 4 pink: 1 yellow ratio should look longer at trials in which 4 yellow and 1 pink balls were sampled than trials in which 4 pink and 1 yellow balls were sampled. Conversely, infants who saw the experimenter sampling from the 4 yellow: 1 pink population should show the opposite looking pattern.

Results

A second observer, blind to trial order, coded 50% of the infants offline. Interscorer reliability averaged 92%. Preliminary analyses found no effects of gender, test trial order (probable-outcome vs. improbable-outcome first) or the population box sampled from (mostly pink or mostly yellow). Subsequent analyses collapsed over these variables.

Looking times for test trial outcomes were analyzed using a 2 x 2 ANOVA with outcome (probable vs. improbable) as the within-subjects factor and age (4-month-olds vs. 6-month-olds) as the between-subjects factor. A significant interaction between Outcome and Age was found, F (1, 30) =7.03, p = .013, effect size (ηp^2) = .190. No other significant main effects or interactions were found.

To break down the interaction, we conducted follow-up t-tests exploring the effect of test trial outcome (probable vs. improbable) for each age group separately (see Table 1 for mean looking times). Six-month-old infants looked reliably longer at the improbable outcome (*M*= 8.63s, *SD*=5.05) than the probable outcome (*M*= 5.96s, *SD*= 2.81), $t(15)$ = 2.67, p = .011. Twelve of sixteen infants looked longer on average at the improbable outcome, Wilcoxon signed ranks test, z = 2.23, p = .013. In contrast, 4-month-olds looked about equally at the improbable outcome (*M* = 6.05s, *SD* = 3.14) and the probable outcome (*M* = 7.45s, *SD* = 5.18), $t(15)$ = 1.19, p =. 250. Seven of sixteen infants looked longer on average at the improbable outcome, Wilcoxon signed ranks test, z = 0.27, p = .605.

Table 1: Mean Looking Times in Seconds By Age Group

	Probable sample (*SD*)	Improbable sample (*SD*)
4-month-olds	7.45 (5.18)	6.05 (3.14)
6-month-olds	5.96 (2.81)	8.63 (5.05)

We also performed a preliminary analysis to address the potential concern that infants only directed their attention to the sampled box and not the additional complementary box despite efforts to draw attention to both population boxes on test trials. Although a more in-depth analysis should include overall duration of looks to each box, to begin we obtained an average number of looks to the sampled and non-sampled boxes for six randomly chosen infants of each age group. We analyzed the looking behavior only during test trials (not familiarization trials) and included infants' looking behavior while the experimenter was sampling and while the static displays were presented after the experimenter revealed the populations. Four-month-old infants looked an average of 7.67 times (*SD* = 6.31) to the sampled box and an average of 10.50 times (*SD* = 8.29) to the non-sampled box. Six month-old infants looked an average of 8.50 times (*SD* = 4.67) to the sampled box and 11.00 times (*SD* = 6.42) to the non-sampled box. There were no differences between the average number of looks to the sampled box between 4-month-olds and 6-month-olds ($t(10)$ = .66, p = .52) or the average number of looks to the non-sampled box between the two age groups ($t(10)$ = .12, p = .91). There were no differences in the number of looks to the sampled vs. non-sampled box for 6-month-olds ($t(10)$ = .77, p = .46) or the number of looks to the sampled vs. non-sampled box for the 4-month-olds ($t(10)$ = .67, p = .52).

Discussion

We found that 6-month-old infants can make generalizations from samples to populations in a probabilistic inference task. When perceptual features are equated in the displays and infants cannot react based solely on perceptual mismatches, 6-month-old infants look longer at the less probable sample of, for example, 4 yellow and 1 pink ball drawn from a mostly pink box than at a more probable sample of 4 pink and 1 yellow ball. Four-month-olds, however, did not show this pattern; they looked roughly equally at both samples. This suggests that the ability to reason about random sampling emerges at around 6 months of age.

One might question whether the current attempt to equate perceptual features in the displays was adequate to rule out the perceptual mismatch account. It is possible that, despite efforts to draw attention to both boxes on stage, infants still only attended to the sampled box and then simply reacted to the mismatch between that box and the sample container on the trials that were probabilistically less likely. In order to assuage this concern, an analysis of infant scanning behavior was performed. The analysis provided preliminary evidence that infants of both ages did attend to both boxes throughout the test trials and they looked a roughly equal number of times at both boxes. This weakens the argument that 6-month-old infants were simply reacting to perceptual mismatches, as 4-month-olds appeared to attend to the same perceptual information as 6-month-olds but did not show differences in looking time between the probable and improbable events. In future work we will include a more detailed analysis of the looking behaviors, providing duration of looks to both boxes and the sample container for each age group.

The findings reported here, in combination with recent evidence from a number of other experimental paradigms, provide strong evidence for early competence in probabilistic reasoning in young infants. In one study, Teglas et al. (2007) found that 12-month-old infants can correctly reason about single-event probability in a looking time task. In this task infants were shown a lottery machine on a computer screen with 3 identical objects and 1 different object moving around in the machine. Infants were able to reason that the 1 different object was more likely to exit the machine on a random draw than 1 of the 3 identical objects. In another study, Denison and Xu (2010b) used a choice task to explore infants' ability to make predictions about single-event probability. In this task, infants were shown two large populations of 4:1 desirable to undesirable objects and 4:1 undesirable to desirable objects. The experimenter removed a single occluded object from each population and infants were able to identify and retrieve the hidden object that was sampled from the population most likely to yield a desirable object (e.g., if infants preferred a pink object, they crawled/walked to the cup that contained an object drawn from the population jar that was mostly pink). Finally, in two studies using the Xu and Garcia (2008) probabilistic inference task, 11-month-old infants have demonstrated the ability to integrate physical and psychological constraints in their statistical computations (Denison & Xu, 2010a; Xu & Denison, 2009). This evidence, from both looking-time and action-based measures, is particularly impressive given the extensive experimental findings suggesting that adults often make faulty probabilistic inferences in a wide range of experimental tasks (e.g., Tversky & Kahneman, 1974, 1981). Kahneman and Tversky find that adult judgments are often hindered by the incorrect application of reasoning heuristics when making probabilistic inferences. The infant findings suggest that humans do have an intuitive, implicit probabilistic reasoning mechanism.

Although we now have evidence of intuitive probabilistic reasoning in 6-month-olds, it appears that 4-month-olds do not share similar intuitions. Two classes of explanations might account for this difference. One possibility is that our task was simply too challenging for 4-month-olds and that competence could be revealed in an even simpler task. For example, they may have had difficulty tracking the sampling behavior of the experimenter, or struggled to encode displays with such a large number of balls. On the other hand, encoding of the displays should be simple, as infants were familiarized to the boxes on a number of trials and the configuration of the boxes did not change throughout the experiment. We will continue to investigate this possibility in future work. A second alternative is that the 4-month-olds' failure could represent a true inability to make generalizations from samples to populations.

If 4-month-olds do in fact lack this ability, it raises the question of why and how it emerges at roughly 6 months. Although we do not wish to commit to any strong claims here, some speculations are in order. First, it is possible that infants learn through experience that samples tend to reflect populations. Between 4 and 6 months of age, infants may accumulate evidence pointing to a relationship between the samples and populations they encounter. For example, they may notice that, day after day, when their mom gives them items from a particular container, the items tend to be tasty. After a number of exposures of this nature, they might eventually make the inference that most items in this container are tasty and thus the container might be filled with a larger quantity of the sampled items. Of course, this requires an initial ability to track regularities in the behaviors of the people and objects in the infant's environment. Once infants gain enough evidence of the predictive relationship between samples and populations, they may begin to use their experience to make generalizations in novel situations such as our task. It could be the case that infants assume random sampling (i.e., strong sampling, see Tenenbaum & Griffiths, 2001 and Xu & Tenenbaum, 2007) and that this is the basis for making inferences given small samples. A second possibility is that this probabilistic reasoning ability emerges through maturation between 4 and 6 months.

Should the capacity to make basic generalizations from samples to populations not emerge until 6 months, we must consider the implications of this suggested developmental time course. If this is a general-purpose mechanism to be used across the lifespan to acquire domain knowledge, how does one reconcile this time of onset with the evidence of genuine domain knowledge – mostly in the domain of physical knowledge – in infants 4 months of age and younger? Again here we can only speculate: One possibility is that the core cognition thesis is correct and much of the domain knowledge present very early on is innately given (Carey, 2009). Another possibility builds on Baillargeon's (2008) suggestion that much early physical knowledge is in fact learned. It is possible that infants are endowed with a very simple object concept such as persistence, and they build on the concept by making revisions based on observed evidence. If domain knowledge, at least regarding physical objects, is acquired in this way, a domain-general mechanism like the one reported here might be at play in informing these revisions early on.

We will attempt to provide further support for the probabilistic account of 6-month-olds' performance in future work. For example, we may test 6-month-olds' abilities to make inferences in the reverse direction of that reported here — from populations to samples (see Xu & Garcia, 2008 Expt. 2 for design details). In this design, infants are required to make probabilistic inferences when they cannot simultaneously view the sample and population. That is, infants will first view the population, the experimenter will occlude the population and a sample will be drawn. During timing, the contents of the population box will no longer be present. This design will help to further rule out the perceptual mismatch account of our findings.

Future experiments will also investigate the sophistication of 6-month-old infants' probabilistic inferences. We will explore whether infants at this age are using the simple heuristic that samples and populations ought to match or if they have a more nuanced conception of probability. Eleven-month-old infants can integrate domain knowledge regarding objects and agents into these probabilistic inference tasks (Denison & Xu, 2010; Xu & Denison, 2009). This integration requires a realization that imposing physical constraints on objects or psychological constraints on agents can drastically change the consequences of sampling. Do 6-month-old infants who appear to be just beginning to show competence in probabilistic inference understand that the consequences of sampling behavior vary depending on sampling conditions? Or, do infants initially deploy this mechanism in all sampling contexts and only later learn the appropriate applications?

Our findings suggest that six- but not four-month-old infants are able to make accurate generalizations from samples to populations and their success in our task cannot be explained by positing a low-level perceptual matching strategy. This pattern of findings suggests that reasoning about the predictive relationship between samples and populations emerges between 4 and 6 months of age.

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. (2009) *The origin of concepts.* Oxford University Press.

Casstevens, R.M. (2007). jHab: Java Habituation Software (Version 1.0.0) [Computer software]. Chevy Chase, MD.

Dawson, C., & Gerken, L. A. (2009). From domain-generality to domain-sensitivity: 4-month-olds learn an abstract repetition rule in music that 7-month-olds do not. *Cognition, 111(3)*, 378-382.

Denison, S. & Xu, F. (2010a). Integrating physical constraints in statistical inference by 11-month-old infants. *Cognitive Science, 34*, 885-908.

Denison, S. & Xu, F. (2010b). Twelve- to 14-month-old infants can predict single- event probability with large set sizes. *Developmental Science, 13*, 798-803.

Gweon, H., Tenenbaum, J.B., & Schulz, L.E. (2010). Infants consider both the sample and the sampling process in inductive generalization. *Proceedings of the National Academy of Sciences, 107(20),* 9066-9071.

Kirkham, N., Slemmer, J., & Johnson, S. (2002). Visual statistical learning in infancy: evidence of a domain general learning mechanism. *Cognition, 83*, B35-B42.

Maye, J., Werker, J., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition, 82(3),* B101-B111.

Schulz, L.E., Bonawitz, E. B., & Griffiths, T. (2007). Can being scared make your tummy ache? Naive theories, ambiguous evidence and preschoolers' causal inferences. *Developmental Psychology 43(5),* 1124-1139.

Tenenbaum, J. & Griffiths, T. (2001). Generalization, similarity, and Bayesian inference. *Behavioral and Brain Sciences, 24*, 629-641.

Teglas, E., Girotto, V., Gonzalez, M., Bonatti, L. (2007). Intuitions of probabilities shape expectations about the future at 12 months and beyond. *Proceedings of the National Academy of Sciences of the United States of America, 104,* 19156-19159.

Tversky, A. & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science 185(4157),* 1124-1131.

Tversky, A.; Kahneman, D. (1981). "The framing of decisions and the psychology of choice". *Science 211(4481),* 453–458.

Xu, F. & Denison, S. (2009). Statistical inference and sensitivity to sampling in 11-month-old infants. *Cognition*, *112,* 97-104.

Xu, F. & Garcia, V. (2008). Intuitive statistics by 8-month-old infants. *Proceedings of the National Academy of Sciences of the United States of America, 105,* 5012-5015.

Xu, F. & Tenenbaum, J. (2007). Sensitivity to sampling in Bayesian word learning. *Developmental Science, 10(3),* 288-297.